AUTHOR QUERY FORM

ELSEVIER	**Journal: GENE** **Article Number: 36807**	**Please e-mail or fax your responses and any corrections to:** **E-mail: corrections.essd@elsevier.spitech.com** **Fax: +1 619 699 6721**

Dear Author,

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Figures 1, 3 and 4 contain text below minimum standard text size. Please check and provide replacement if deemed necessary.
Q2	Please supply the year of publication.
Q3	Please supply the year of publication.
Q4	The sub-panel (A) is found in Figure 1, 2 artwork. However, no sub-panels (B) and (C) is found in the artwork. Kindly check and correct if necessary.

Thank you for your assistance.

Supplementary material.

Gene xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system

Vivek Anantharaman, Saraswathi Abhiman, Robson F. de Souza, L. Aravind *

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

ARTICLE INFO

Article history:
Accepted 6 December 2010
Available online xxxx

Received by I.B. Rogozin

Keywords:
Protein sequence analysis
GPCR
RGS
G-proteins
START domain
PAS domain
CHASE domain
7TM

ABSTRACT

Though the heterotrimeric G-proteins signaling system is one of the best studied in eukaryotes, its provenance and its prevalence outside of model eukaryotes remains poorly understood. We utilized the wealth of sequence data from recently sequenced eukaryotic genomes to uncover robust G-protein signaling systems in several poorly studied eukaryotic lineages such as the parabasalids, heteroloboseans and stramenopiles. This indicated that the Gα subunit is likely to have separated from the ARF-like GTPases prior to the last eukaryotic common ancestor. We systematically identified the structure and sequence features associated with this divergence and found that most of the neomorphic positions in Gα form a ring of residues centered on the nucleotide binding site, several of which are likely to be critical for interactions with the RGS domain for its GAP function. We also present evidence that in some of the potentially early branching eukaryotic lineages, like *Trichomonas*, Gα is likely to function independently of the Gβγ subunits. We were able to identify previously unknown Gγ subunits in *Naegleria*, suggesting that the trimeric version was already present by the time of the divergence of the heteroloboseans from the remaining eukaryotes. Evolution of Gα subunits is dominated by several independent lineage-specific expansions (LSEs). In most of these cases there are concomitant, independent LSEs of RGS proteins along with an extraordinary diversification of their domain architectures. The diversity of RGS domains from *Naegleria* in particular, which has the largest complement of Gα and RGS proteins for any eukaryote, provides new insights into RGS function and evolution. We uncovered a new class of soluble ligand receptors of bacterial origin with RGS domains and an extraordinary diversity of membrane-linked, redox-associated, adhesion-dependent and small molecule-induced G-protein signaling networks that evolved in early-branching eukaryotes, independently of parallel systems in animals. Furthermore, this newly characterized diversity of RGS domains helps in defining their ancestral conserved interfaces with Gα and also those interfaces that are prone to extensive lineage-specific diversification and are thereby responsible for selectivity in Gα–RGS interactions. Several mushrooms show LSEs of Gαs but not of RGS proteins pointing to the probable differentiation of Gαs in conjunction with mating-type diversity. When combined with the characterization of the 7TM receptors (GPCRs), it becomes apparent that, through much of eukaryotic evolution, cells contained both 7TM receptors that acted as GEFs and those as GAPs (with C-terminal RGS domains) for Gαs. Only in some lineages like animals and stramenopiles the 7TM receptors were restricted to GEF only roles, probably due to selection imposed by the rate-constants of the Gαs that underwent lineage-specific expansion in them. In the alveolate lineage the 7TM receptors occur independently of heterotrimeric G-proteins, suggesting the prevalence of G-protein-independent signaling in these organisms.

Published by Elsevier B.V.

1. Introduction

The regulatory role of heterotrimeric G-proteins was one of the first signaling mechanisms to be studied in depth in eukaryotes and represents a classical paradigm in signal transduction (Rodbell et al., 1971; Cassel and Selinger, 1976; Ross and Gilman, 1977; Abramowitz et al., 1979; Freissmuth et al., 1989). In animals heterotrimeric G-proteins are central to the mechanisms which transmit intracellular signals in response to key sensory signals such as light (vision), extrinsic chemicals (taste and smell), internal physiological states (neurotransmitters, hormones) and immunity-related cues (e.g. chemokines) (Freissmuth et al., 1989; Neves et al., 2002; Bastiani and Mendel, 2006; Wensel, 2008; Cho and Kehrl, 2009). Molecular studies have revealed that the common denominator in these signaling processes is a group of proteins that engage in dynamic interactions with each other. These include: 1) a receptor with 7 transmembrane segments (also called GPCRs or serpentine receptors; hereinafter 7TM receptors),

Abbreviations: GPCRs, C-protein coupled receptors; 7TM, 7-transmembrane receptor; GAP, GTPase activating protein; PBP-I, periplasmic-binding protein type-I domain; TRAFAC GTPase, translation factor class; GAP, GTPase-activating protein; RGS, regulator of G-protein signaling; GEF, GDP exchange factor; LSE, lineage-specific expansion.

* Corresponding author. Tel.: +1 301 594 2445.
E-mail address: aravind@mail.nih.gov (L. Aravind).

0378-1119/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.gene.2010.12.001

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

UNCORRECTED PROOF

ARTICLE IN PRESS

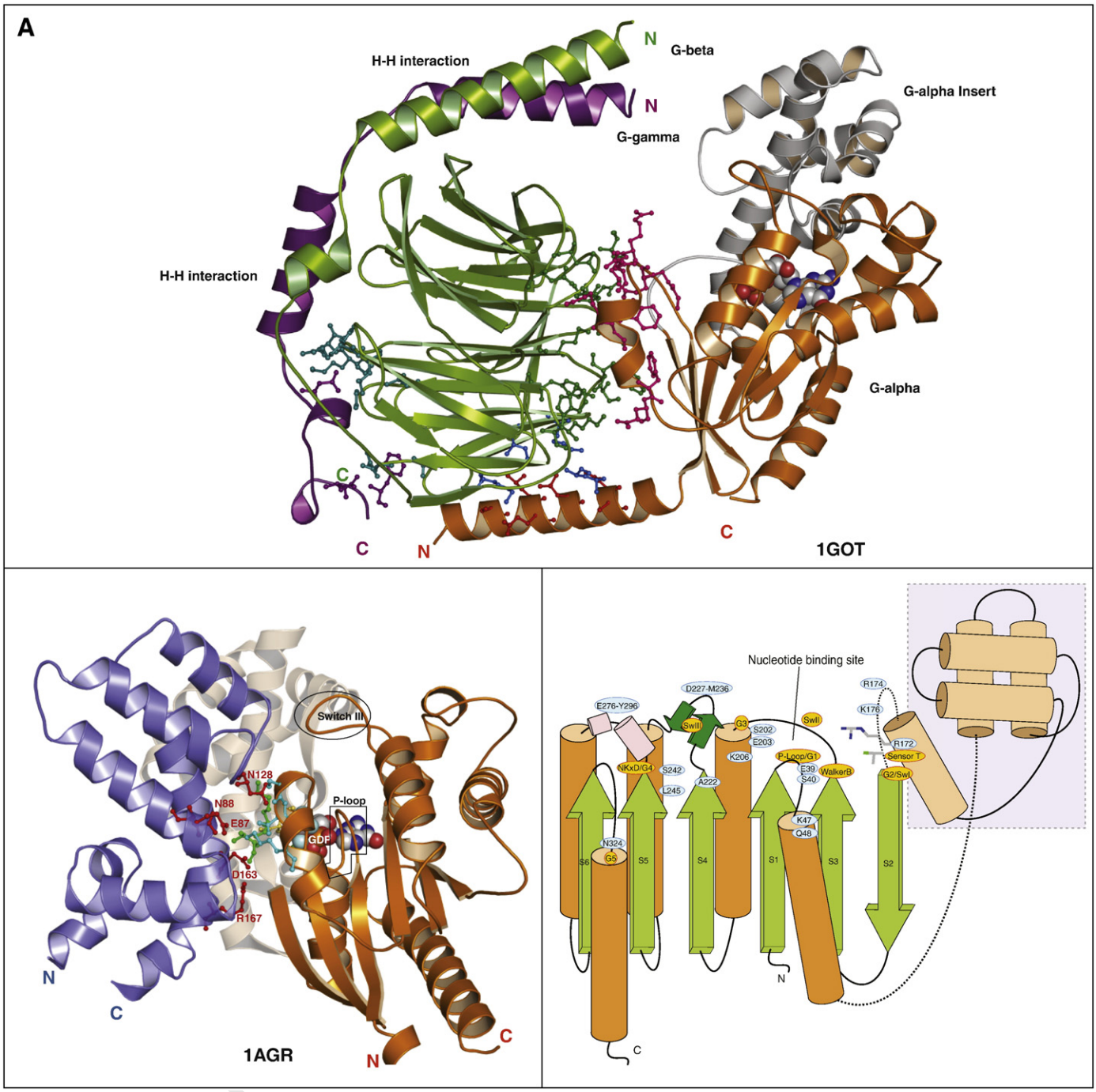
which might receive the sensory input by means of a bound prosthetic group (e.g. sensing of light by the retinaldehyde group in rhodopsin), or by direct interaction with the sensed molecule (e.g. in odorant receptors), or by means of a linked N-terminal extracellular domain that binds ligands (e.g. the periplasmic-binding protein domain (PBP-I) in the taste receptors) (Pierce et al., 2002; Yeagle and Albert, 2007; Hanson and Stevens, 2009). 2) An intracellular heterotrimeric GTPase with a catalytic subunit Gα that contains a low-efficiency GTPase domain of the extended Ras-like clade of translation factor class (TRAFAC) of GTPases (Leipe et al., 2002). The Gβ subunit is a WD40-type β-propeller protein with 7 blades that interacts with the Gγ subunit, a small domain comprised of two core α-helices (Freissmuth et al., 1989; Neves et al., 2002; Neuwald, 2007). 3) A GTPase-activating protein (GAP) for the Gα subunit, which contains a conserved regulator of G-protein signaling (RGS) domain that facilitates the high turnover of GTP hydrolysis by physically interacting with Gα (Sethakorn et al.; Berman et al., 1996; Siderovski et al., 1996; Popov et al., 1997; Soundararajan et al., 2008; Cho and Kehrl, 2009; McCoy and Hepler, 2009; Porter and Koelle, 2009; Tesmer, 2009).

In animals, the typical signaling cycle by this protein network is initiated by binding of the ligand to or conformational change in the prosthetic group attached to the 7TM. This induces the receptor to act as a GDP exchange factor (GEF) for Gα, exchanging GDP for GTP. The GTP-bound Gα dissociates from the βγ dimer and becomes ready to associate with diverse effectors such as adenylyl cyclases and phospholipases to transmit the signal. Thus, the βγ dimer acts as a GDP dissociation inhibitor for the GTPase (Siderovski and Willard, 2005). Additionally, the βγ dimer might also associate with certain effectors and independently transmit a distinct signal. The signaling switch is completed by the action of the RGS domain protein that activates GTP hydrolysis, returning the Gα to the GDP-bound state in which it reassociates with the βγ dimer (Freissmuth et al., 1989; Neves et al., 2002; Siderovski and Willard, 2005; Bastiani and Mendel, 2006; Wensel, 2008). Structural studies over the past 15 years have revealed the details of this process at an atomic level. Firstly, the structure of Gα confirmed the observation that it contains an in-built arginine finger, which is provided by a unique helical insert just N-terminal to the region of the G2 motif bearing the switch-I threonine (Bourne et al., 1991; Noel et al., 1993; Lambright et al., 1996) (Fig. 1). In this sense the Gαs differ from most other members of the extended Ras-like clade of GTPases, which have the arginine finger provided by an external GAP domain protein (Scheffzek et al., 1998). The structure of the Gαβγ trimer showed that the β subunit fits tightly into a slot formed by the α-helix N-terminal to the core GTPase domain and the GTPase β-sheet in the region of the G2 motif (Lambright et al., 1996). The β subunit has the bi-helical N-terminal extension to the β-propeller, which forms a helix–helix coil with the corresponding bi-helical unit in the γ subunit. Further, the γ subunit has a conserved "squiggle" comprised of a single helical turn and an extended region at the C-terminus, both of which are inserted into pockets formed by the two terminal blades of the β-subunit's propeller (Fig. 1) (Lambright et al., 1996). In most Gαs the N-terminal helix of the α-subunit is processed at the second glycine and modified by a lipid chain that facilitates its association with the membrane (Escriba et al., 2007; Wensel, 2008) (Fig. 2). Likewise, most γ subunits are processed to reveal a C-terminal cysteine, whose side chain is farnesylated or geranylgeranylated and carboxyl group is methylated (Fig. 2) (Escriba et al., 2007; Wensel, 2008). This allows the βγ to associate with the membrane. The structure of the Gα–RGS complex revealed that, unlike GAPs of other Ras-like GTPases, the RGS domain does not contribute an arginine finger, but appears to interact with all the three switch regions and convert the Gα into a high activity GTPase (Tesmer et al., 1997; Soundararajan et al., 2008; Tesmer, 2009). However, certain other proteins such as axin and the G-protein-receptor kinases contain RGS domains that have no GAP activity but merely serve as an interface to interact with G proteins (Sethakorn et al.; Spink et al., 2000; Dajani et al., 2003; Tesmer et al., 2005; Tesmer, 2009). Yet other RGS proteins, such as those found in ARHGEF11, have lost the conventional GTPase interaction interface but have independently reacquired GAP capability via an N-terminal extension (Chen et al., 2005; Chen et al., 2008).

Studies on fungal mating interactions and glucose-sensing in plants revealed that 7TM receptors and heterotrimeric GTPase-RGS comprise the central switch in sex-pheromone and chemoreception signaling, thereby extending the phylogenetic generality of this paradigm (De Vries et al., 1995; Druey et al., 1996; Koelle and Horvitz, 1996; Siderovski et al., 1996; Versele et al., 2001; Yu, 2006). Subsequently, studies in plants have shown that a 7TM-G-protein network is required for the sensing and signaling the presence of sugars such as glucose (Jones, 2002; Chen et al., 2003; Siderovski and Willard, 2005; Johnston et al., 2007; Grigston et al., 2008; Johnston et al., 2008). Interestingly, one of the plant 7TM receptors, RGS1, in this sugar sensory network contains a C-terminal RGS domain in its cytoplasmic tail. This RGS domain was found to be required for active signaling by maintaining a high level of GTPase activity, which was found to be the rate limiting step, as against the animal models where the GDP-GTP exchange was found to be rate limiting (Johnston et al., 2007; Grigston et al., 2008; Johnston et al., 2008). Plants were also found to encode a class of divergent Gαs (the XLG proteins) that were claimed to bind GTP, though there is no evidence for their association with 7TM receptors (Pandey et al., 2008; Zhu et al., 2009). These results in plants indicated for the first time that alternatives to the mechanistic themes found in the animals might be mediated by the same set of conserved domains in other eukaryotes. Likewise, studies in another model system, *Dictyostelium*, have revealed a wide range of RGS domains that might mediate unique signaling processes downstream of the multiple Gαs found in this organism (Natarajan et al., 2000; Brzostowski and Kimmel, 2001; Janetopoulos et al., 2001; Prabhu and Eichinger, 2006; Elzie et al., 2009). For example, the protein RCK1 has been shown to be a negative regulator of chemotaxis in *Dictyostelium* downstream of the cNMP-sensing 7TM receptor and combines a RGS domain with a C-terminal kinase domain (Sun and Firtel, 2003). In this case it has been shown that the disassociation of the RGS protein from the receptor complex at the membrane requires the activity of the kinase. Newly sequenced eukaryotic genomes (e.g. the potentially early-branching eukaryote *Trichomonas vaginalis*) were also found to encode proteins similar to plant 7TM receptors with RGS domains (Fig. 3) (Fritz-Laylin et al.; Johnston et al., 2008). These findings from comparative genomics suggest that heterotrimeric G-protein signaling has a major presence outside of the well-studied model organisms such as plants, fungi and animals. These observations, together with the indications of mechanistic differences in heterotrimeric G-protein signaling in plants, suggested that a comprehensive comparative genomics study of this regulatory network across eukaryotes might throw new light on its functions and evolution.

In the current article we perform a comprehensive comparative genomic survey and sequence-structure analysis of the heterotrimeric G-protein signaling network, including the three subunits of the GTPase complex, the RGS domain proteins and the receptors feeding into them. Based on these analyses we attempt to address several problems that include: 1) The provenance and evolution of the individual components and their interactions. 2) Identification of sequence and structure determinants that allow robust extension of the extensive understanding of this pathway that has accrued in model systems (mainly animals and fungi and to a lesser degree plants) to non-model eukaryotes. 3) Inference of novel paradigms of heterotrimeric G-protein signaling in non-model eukaryotes and their possible implications for model systems. 4) Understanding and measuring the overall diversity of this signaling network in course of eukaryotic evolution. As a consequence of the investigation of the above issues we made several novel observations that we synthesize to elucidate several aspects of the natural history of the G-protein-7TM receptor-based signaling system.

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

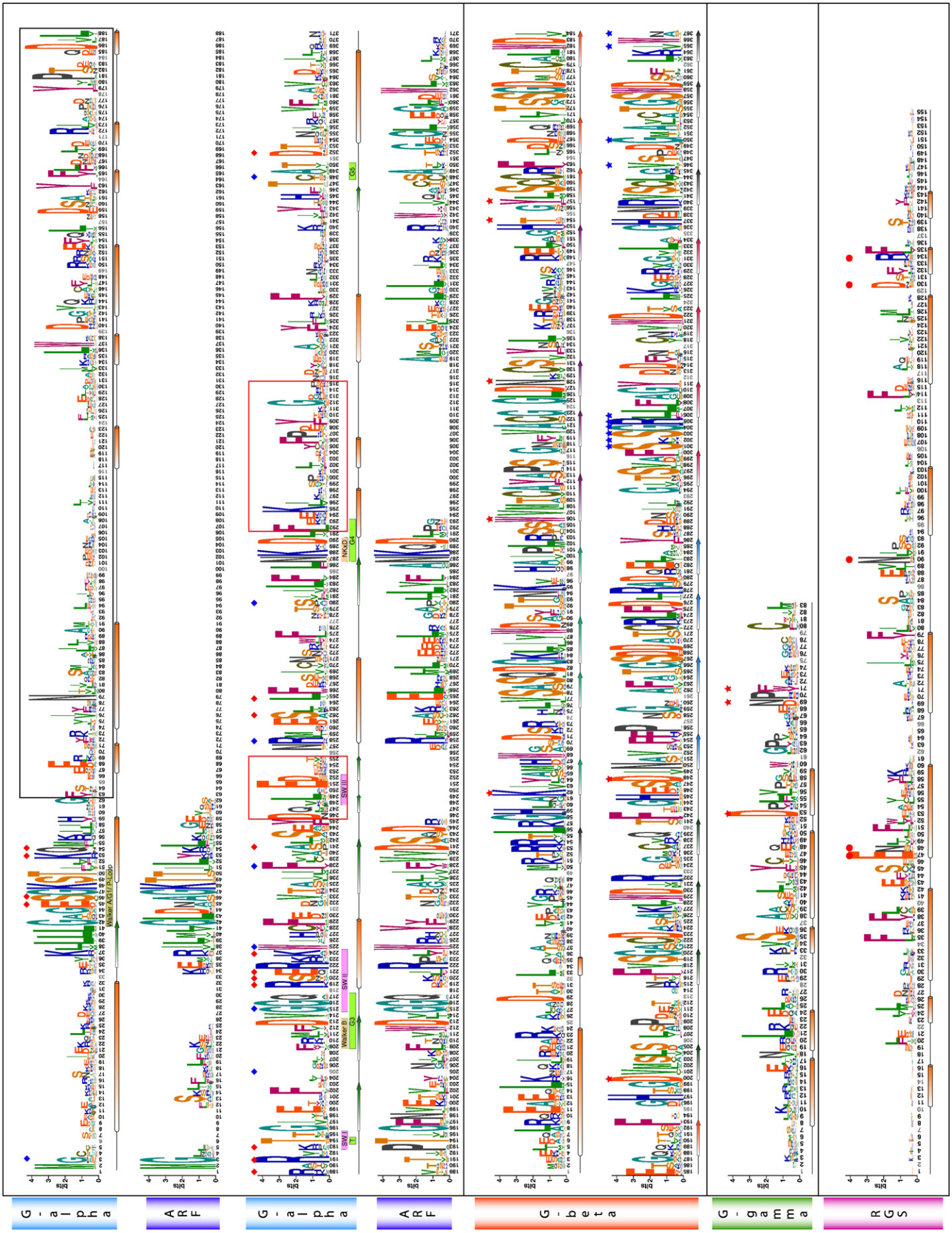


Q1 **Fig. 1.** (A) A cartoon representation of the heterotrimeric G-protein (PDB: 1got). Gα is colored in orange, whereas the α-helical insert is shown in grey color; Gβ is colored in green and Gγ in purple. The substrate GDP is shown as spheres. Interacting residues are colored as follows: red: residues in Gα (N-term helix) interacting with Gβ; Blue: residues in Gβ interacting with Gα (N-term helix); magenta: residues in Gα interacting with Gβ; green: residues in Gβ interacting with Gα; purple: residues in Gγ interacting with Gβ; teal: residues in Gβ interacting with Gγ. (B) Gα with RGS. A cartoon representation of the Gi-RGS4 complex (pdb: 1agr) is shown. Gα is colored in orange, with the insert shown in grey color and RGS is shown in blue color. Residues in RGS interacting with Gα are shown in red color. Interacting residues in Gα are shown in green and cyan colors. P-loop is marked by a box. N-terminal helix of Gα is only shown partially. (C) A topological representation of Gα is shown. The active site motifs are labeled in yellow. The 45 conserved positions that are unique to Gα when compared with the ARFs are labeled in blue ovals. The numberings correspond to the "idealized" Gα (hybrid of transducin and Gi) represented by pdb: 1got. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Materials and methods

Iterative profile searches with PSI-BLAST (Altschul et al., 1997) were used to retrieve homologous sequences in the protein non-redundant (NR) database at National Center for Biotechnology Information (NCBI). A profile inclusion expectation (E) value threshold of 0.01, the BLOSUM62 substitution matrix and the gap penalty (existence: 11 and extension: 1) were utilized in the searches. Similarity-based clustering for both classification and culling of nearly identical sequences was performed using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.txt). Hidden Markov model searches were performed using the HMMSEARCH (for HMM built from an alignment) and JACKHMMER (for single query sequence) programs of the HMMer 3.0 package (http://hmmer.janelia.org/). Multiple sequence alignments were built using Kalign (Lassmann and Sonnhammer, 2005) and PCMA (Pei et al., 2003) followed by manual

ARTICLE IN PRESS



Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

ARTICLE IN PRESS

adjustments based on profile-profile alignment, secondary structure information and structural alignment. Consensus secondary structures were predicted using the JPred program (Cuff and Barton, 2000). Multiple alignments of individual domains discussed here are provided in the supplementary material. Profile-profile comparisons were performed using the HHpred program (Soding et al., 2005). HHpred utilizes a comparison between HMM constructed using the query via a PSI-BLAST search and a preexisting library of HMM compiled from domains found in the PDB database to detect remote relationships. Structure similarity searches and structural alignments were performed using the DaliLite program that searches the PDB database with coordinates of a query structure and makes structural alignments (Holm and Sander, 1998). For previously characterized domains the PFAM database was used as a guide along with in-house PSI-BLAST profiles for each of these domains. Clustering with BLASTCLUST followed by multiple sequence alignment and further sequence profile searches were used to identify other domains that were not present in the PFAM database.

Entropy calculations were performed with a custom script using the alignments generated by KALIGN (Lassmann and Sonnhammer, 2005). The entropy was calculated using the Shannon entropy formula: $= -\sum_{i=1}^{M} Pi \log 2Pi$, where Pi is the fraction of a given amino acid i, and M the total number of different amino acids. Structural manipulations were carried out using the Swiss-PDB viewer and PYMOL programs (Guex and Peitsch, 1997; DeLano, 2002). Unrooted phylogenetic trees reconstructed using an approximately-maximum-likelihood method implemented in the FastTree 2.1 program under default parameters (Price et al., 2010) and alternatively using the MEGA program (Kumar et al., 2001). In contrast to other phylogenetic analysis programs, FastTree stores sequence profiles of internal nodes. This improves speed considerably for large alignments like those used in this study and in our benchmarking efforts it provides trees comparable to those generated with the Protml program initiated with starting trees constructed with neighbor-joining or minimal evolution methods. The domain architecture graphs were rendered using Cytoscape (Cline et al., 2007). Signal peptides were predicted using the SIGNALP program (Nielsen et al., 1997). Multiple alignments of the N-terminal regions of proteins were used additionally to verify the presence of a conserved signal peptide, and only those signal peptides that were conserved across the aligned group of proteins were considered as true positives. Transmembrane regions were predicted in individual proteins using the TMHMM2.0 and Phobius programs with default parameters (Krogh et al., 2001; Kall et al., 2004).

3. Results and discussion

3.1. Identification of heterotrimeric G-protein subunits, RGS domain proteins and 7TM receptors and inference of the phyletic profiles

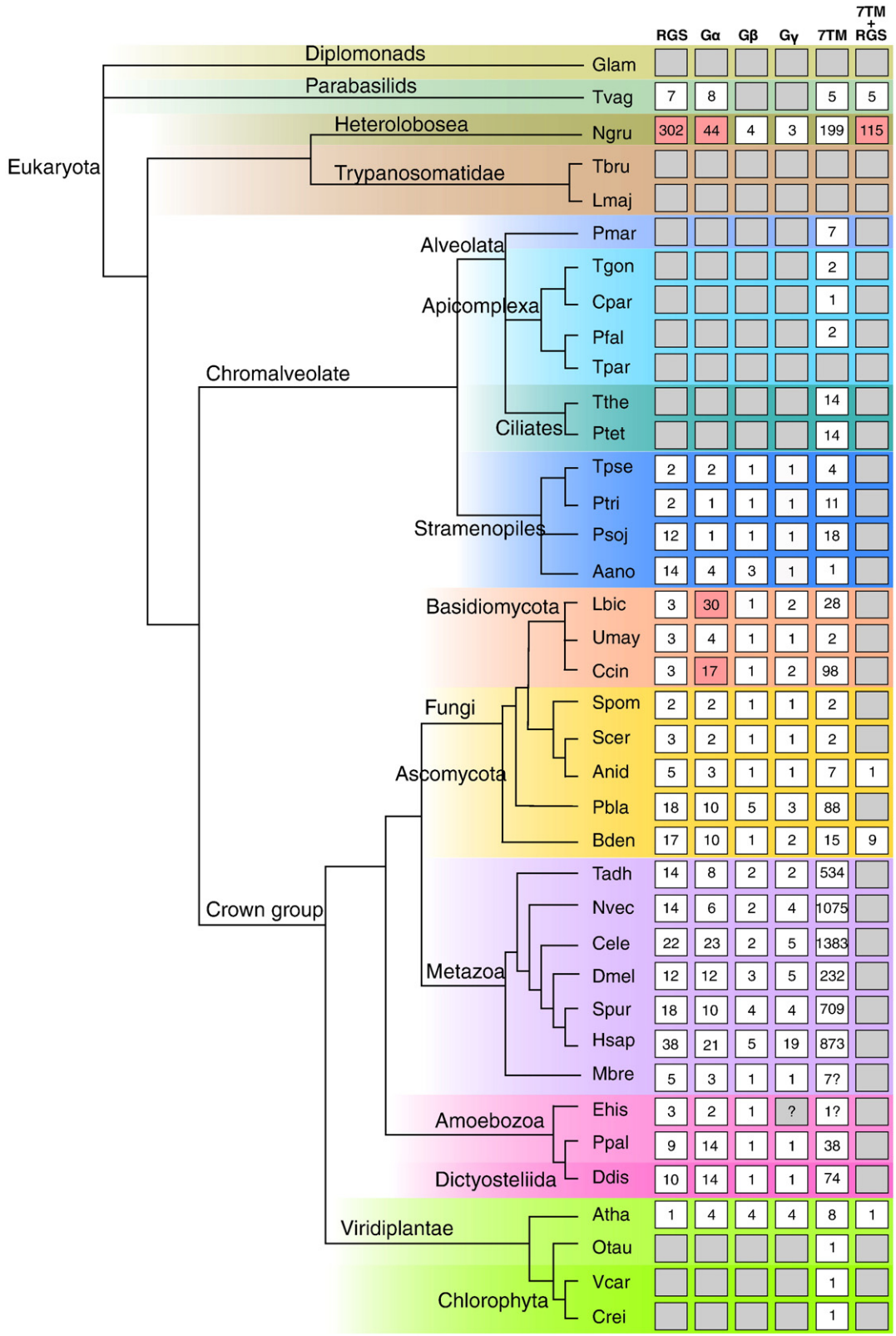
By combining sequence profile searches with domain architecture and structural analysis (see Methods for details) we were able to reliably identify representatives of all the three subunits of heterotrimeric G-proteins, RGS domains and 7TM receptors (Fig. 3, supplementary material). Given that 7TM receptors can be extremely divergent we also used a second strategy of identifying all TM proteins in a given proteome using TMHMM and Phobius and extracting those with 7 predicted TM segments (Anantharaman and Aravind, 2003). These were then clustered using the BLASTCLUST program and used as seeds for further searches using the PSI-BLAST and Jackhmmer programs. We then prepared alignments for all positively identified complete heterotrimeric G-protein subunits and the RGS domains and used this alignment to infer column-wise Shannon entropy for each of the alignments (Supplementary material). Along with the sequence alignments we also prepared structural alignments of PDB coordinates of each of the components with the DALIlite program and used these as templates to further combine them with the sequence-based alignments. We also used structures of various G-protein complexes in conjunction with the above alignments to perform a systematic structure-function analysis of each of the component domains of the system and clarify their potential mode of action. Both similarity-based clustering and phylogenetic analysis were used to classify them further and infer the final phyletic profiles for each of the components of the signaling system under consideration. We briefly outline below specific features of each of the components that emerged as a result of this analysis.

3.2. The origin and diversification of the Gα subunit

The Gα subunit belongs to the clade of extended Ras-like GTPases within the great TRAFAC class of GTPases (Leipe et al., 2002). The extended Ras-like clade was previously shown to be divided into two major groups: 1) The MglA-Arf-Gα group and 2) The Ran–Ras–Rho–Rab-like group (Leipe et al., 2002; Jekely, 2003; Neuwald, 2007). Given that versions belonging to both the major groups are found in archaea and bacteria, this fundamental split within the extended Ras-like clade occurred within the prokaryotes, prior to the origin of eukaryotes (Leipe et al., 2002; Dong et al., 2007). The eukaryotes appear to have acquired an MglA-like protein from their archaeal precursor whereas the precursor of the second major group could have come from either the archaeal progenitor or the primary bacterial endosymbiont (Koonin and Aravind, 2000; Leipe et al., 2002; Dong et al., 2007). Founders of each of the two major groups appear to have proliferated into several distinct families very early in eukaryotic evolution in relation to the diversification of the internal membranes of the eukaryotic cell (Jekely, 2003). The eukaryotic Arf-like (including the Sar subfamily) proteins arose from the MglA-like progenitor and acquired a primary function related to protein trafficking both in the context of the endoplasmic reticulum and the Golgi apparatus (Leipe et al., 2002; Jekely, 2003; Kahn, 2009). The Arf-like proteins retained most of the ancestral features of the extended Ras-like clade of GTPases, such as the small size without any inserts in the core GTPase domain, but acquired a distinct N-terminal helix with a conserved glycine which provided the site for the lipid anchor (Fig. 2) (Escriba et al., 2007; Kahn, 2009). In contrast, in the Ran-Ras-Rho-Rab-like group the lipid anchor emerged in the form of a C-terminal extension with a single or a pair of conserved cysteines. In addition to the earlier noted elements of similarity between Gα and the Arf-like proteins, such as the shared signatures spanning the Walker B (G3) motif and switch-2 region (Neuwald, 2007) (Supplementary material), we identified several additional features that support the above conjecture (Fig. 2 and Supplementary material): 1) N-terminal helical extension with a glycine for lipid modification; 2) A shared arginine immediately downstream of the switch-3 region; 3) a shared deletion of 4-5 amino acids relative to the other Ras-like GTPases just

Fig. 2. The sequence logos were made using multiple sequence alignments of respective domains (Crooks et al., 2004). Consecutive gap-rich positions in the alignment are reduced to a single gap marked by grey numbers in the logo. Secondary structures are represented at the bottom of the logos. Various conserved motifs are marked with boxes below the alignment position numbers. (A) Gα vs Arf: the red diamonds are conserved positions that are unique to Gα when compared with the ARFs. These are the same positions that are marked in Fig. 1. The blue diamonds indicate positions which are synapomorphies of Gαs and ARFs. The black box marks the alpha-helical insert in Gα and the red boxes mark the Switch III insert and the bi-helical insert after S5 in Gα. T is the "sensor threonine" and the G2 motif is centered on this residue. (B) Gβ: conserved residues that interact with Gα are shown with a red star on top of the sequence logo. Blue stars on top of the logo mark the conserved residues that interact with Gγ. The propellers are shown using discretely colored strands, with the first strand colored same as the last propeller to show circular permutation. (C) Gγ: conserved residues that interact with Gβ are shown with a red star on top of the sequence logo. (D) RGS: conserved residues that interact with Gα are shown with a red circle on top of the sequence logo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001



upstream of the Walker B motif. 4) Several members of the Arf-like clade and the Gα share a cysteine which replaces the usual serine in the SA signature seen in the G5 motif. These observations suggest that the Gα subunit emerged in eukaryotes from an Arf-like precursor, after the acquisition of the N-terminal helix with the lipid modified glycine, through drastic divergence.

Analysis of the phyletic patterns of Gα suggests a sporadic but widespread presence across eukaryotes (Fig. 3). All members of the eukaryotic clade including the animals, fungi, amoebozoans, and plants, encode 1-30 Gα paralogs (Aubry and Firtel, 1998; Janetopoulos et al., 2001; Bastiani and Mendel, 2006; Johnston et al., 2008; Zhu et al., 2009). Within this clade the lowest numbers are observed in plants, which

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

ARTICLE IN PRESS

typically encode only a single conventional Gα, though they might encode multiple paralogs of divergent XLG type (Fig. 3) (Zhu et al., 2009). Surprisingly, we could not detect any Gαs in the available chlorophyte genomes suggesting that it might have been lost early in the evolution of chlorophytes. Outside of animals and slime molds, which typically have large numbers of Gα paralogs (often >10), we found atypically large complements of Gαs in certain fungi such as mushrooms (e.g. 30 paralogs in *Laccaria*), chythrids and mucors (Fig. 3). They are entirely absent in alveolates but are present in at least a single copy in stramenopiles and haptophytes. A number of phylogenetic studies suggest that the kinetoplastids, heteroloboseans, diplomonads and parabasalids represent eukaryotic lineages that were among the first to branch of from the rest of the eukaryotes (Arisue et al., 2005; Simpson et al., 2006; Iyer et al., 2008). Among these, Gαs are absent in the kinetoplastids and diplomonads, but are highly expanded in *Naegleria* (the largest set among currently sequenced eukaryotic genomes; 44 paralogs) and to a lesser extent in *Trichomonas* (Fig. 3). Phylogenetic analysis indicates that the common ancestor of metazoans had six Gα paralogs that were cognates of the human Gi, Go, G12/13, Gq and Gs type and of the zebrafish Gv type Gαs (Simon et al., 1991; Bastiani and Mendel, 2006; Oka and Korsching, 2009). In contrast the sister-group of the metazoans, the choanoflagellate *Monosiga*, contains only three Gαs (Fig. 3), suggesting that a part of the expansion might have occurred after the metazoans branched off from the former. Of these the Gv type appears to have been repeatedly lost in several metazoan lineages. In *Caenorhabditis* there were at least two distinct lineage specific expansions (LSEs) derived from the Gi/Go clade which appear to be related to the development of the elaborate nematode chemosensory system (Bastiani and Mendel, 2006). The common ancestor of fungi can be reconstructed as possessing three Gα paralogs which had duplicated independently of the five inferred in the metazoan ancestor. Two of these corresponded to Gpa1 and Gpa2 of *Saccharomyces cerevisiae.* The third version underwent independent lineage-specific expansions in both mushrooms and chythrids but appears to have been lost in other fungal lineages. Among the plant versions it is clear that the conventional Gα and the XLG versions emerged through a duplication that happened prior to the divergence of modern land plants including the mosses (the clade embryophyta). The radiations observed in the basal eukaryotes such as *Naegleria* and *Trichomonas* are not specifically related to any of the versions seen in the animals and represent independent LSEs in these organisms. Given the potentially basal position of *Naegleria* and *Trichomonas* it appears plausible that the last eukaryotic common ancestor (LECA) possessed at least one Gα paralog suggesting that it had already diverged from the Arf-like GTPases by that period. Within eukaryotes, the tree suggests that only a single Gα can be inferred in the common ancestor of the clade including the animals, fungi, plants and haptophytes and stramenopiles (Fig. 3), which is consistent with the idea that the evolution of the Gαs was dominated primarily by LSEs or complete loss.

3.3. Structural features that are correlated with function and identification of deviant forms of Gαs

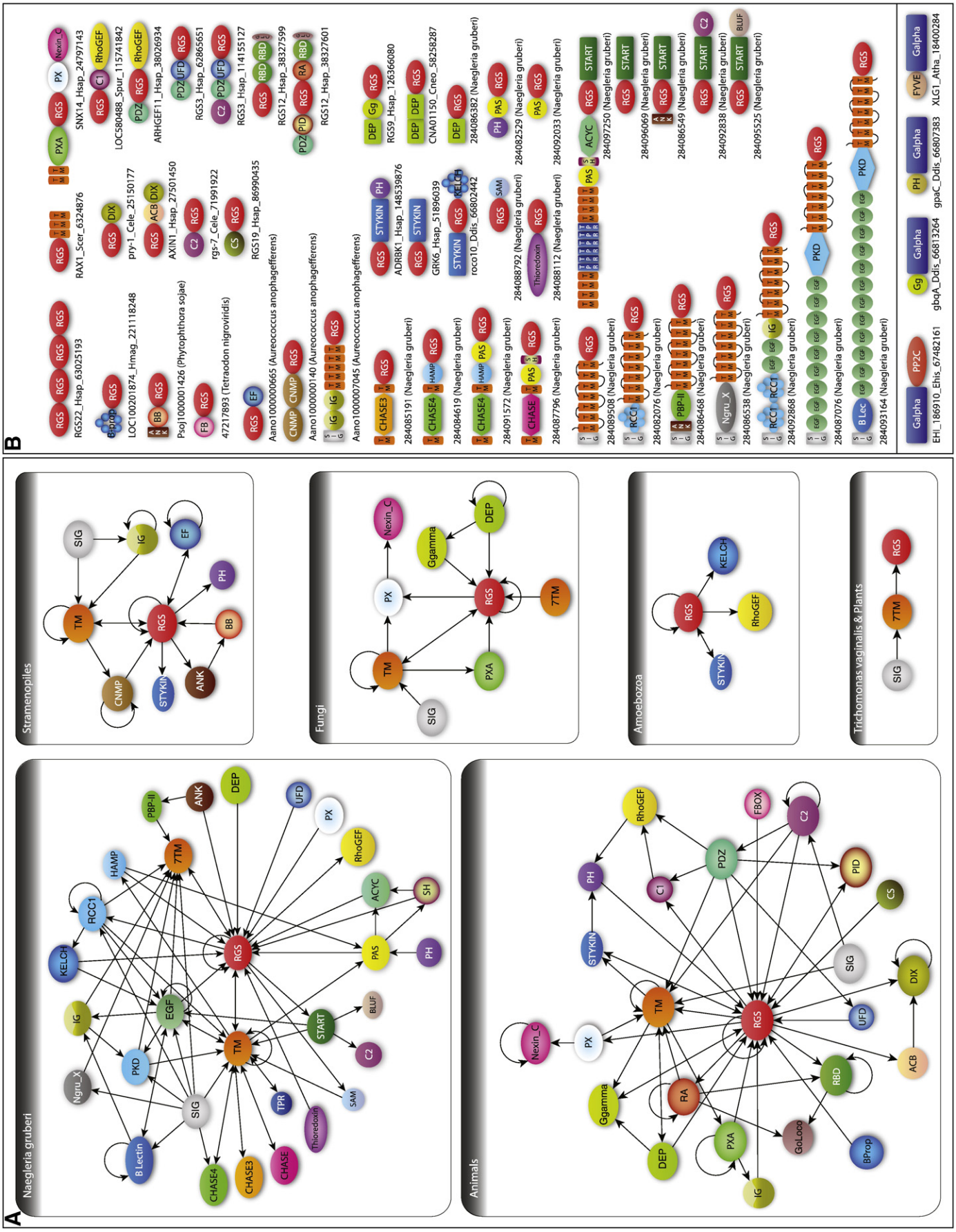
Given that previous studies on Gαs had largely concentrated on representatives from model organisms, we sought to use their diversity from across the eukaryotes to identify novel aspects of their biochemistry and biology. Firstly, we used the more diverse set of Gαs to better infer the conserved features that define them and delineate those that emerged in course of their rapid and drastic divergence from Arf-like GTPases (Neuwald, 2007). The most obvious features that emerged in the Gαs relative to the ancestral Arf-like proteins are the two major inserts (Figs.1, 2): 1) the large α-helical sub-domain inserted between the G1 and G2 motifs, the C-terminal part of which supplies the arginine finger and 2) the switch-3 region, which is a small β-hairpin inserted after the 4th strand of the GTPase domain. An earlier comparison of Gαs with the Arf-like GTPases had shown that the former possess a unique lysine immediately downstream of the switch-2 region that undergoes a conformational change between the GTP- and GDP- bound forms (Neuwald, 2007). This residue was shown to be critical for the differential interaction of these forms with the Gβ subunit. Our comparison (Figs. 1, 2 and Supplementary material) identified several additional positions that are conserved specifically in the Gαs relative to the Arf-like proteins. We plotted these positions (including switch-3, totally 45 positions) on the structure of Gα representatives, the human transducin (Lambright et al., 1996) and Gi (Tesmer et al., 1997) (PDB: 1got and 1agr; Supplementary material). Strikingly, with the exception of three residues downstream of the Switch-2 region that are part of the primary interface with both Gβ and RGS (including the earlier observed lysine (Lambright et al., 1996; Neuwald, 2007)), almost all the remaining positions form a ring around the nucleotide-binding site. The probability that these 42 positions drawn from the total of 343 positions in Gα lie in this ring by chance alone is very low ($p<.001$, by random sampling using the idealized Gα topology), suggesting that they relate to specific functional features that emerged in the Gαs after divergence from the Arf-like proteins. Examination of these residues shows that a subset of them, primarily drawn from the G1 (P-loop) and G2 regions, form a cage around the nucleotide (Figs. 1, 2). Those from the switch-3 region overlap with the set of residues involved in forming an interface with the RGS (Tesmer et al., 1997; Soundararajan et al., 2008), while those from the helical insert downstream of the G4 motif are placed externally forming one of the interfaces for effector interaction (Berlot and Bourne, 1992). Residues in this ring show a network of contacts with each other via hydrogen bonds, hydrophobic and van der Waals interactions suggesting that they could physically communicate conformational changes arising from the state of the bound nucleotide. Comparison of this ring in the GDP-bound form (e.g. PDB: 1got (Lambright et al., 1996)) with the form bound to the GTP or its mimic (e.g. PDB: 1agr (Tesmer et al., 1997; Soundararajan et al., 2008)) shows that it is more open in the former state (Supplementary material). In the later state the interaction of the RGS with the switch-

Fig. 3. The eukaryotic phylogenetic tree is shown with the absolute protein counts for each organism. The expansions are shown in pink and absences are shown in grey. The 7TM counts are only the lower bounds because certain families could have eluded detection due to their great lineage specific diversity. The following organisms were analyzed for this study: Afum: *Aspergillus fumigatus*; Anid: *Aspergillus nidulans*; Cgla: *Candida glabrata*; Calb: *Candida albicans*; Cglo: *Chaetomium globosum*; Dhan: *Debaryomyces hansenii*; Egos: *Ashbya gossypii*; Gzea: *Gibberella zeae*; Klac: *Kluyveromyces lactis*; Mgri: *Magnaporthe oryzae*; Ncra: *Neurospora crassa*; Psti: *Scheffersomyces stipitis*; Pnod: *Phaeosphaeria nodorum*; Scer: *Saccharomyces cerevisiae*; Spom: *Schizosaccharomyces pombe*; Ylip: *Yarrowia lipolytica*; Cneo: *Cryptococcus neoformans*; Ccin: *Coprinopsis cinerea*; Ppla: *Postia placenta*; Umay: *Ustilago maydis*; Lbic: *Laccaria bicolor*; Bden: *Batrachochytrium dendrobatidis*; Ecun: *Encephalitozoon cuniculi*; Ebie: *Enterocytozoon bieneusi*; Pbla: *Phycomyces blakesleeanus*; Amel: *Apis mellifera*; Dmel: *Drosophila melanogaster*; Apis: *Acyrthosiphon pisum*; Dpul: *Daphnia pulex*; Tcas: *Tribolium castaneum*; Nvit: *Nasonia vitripennis*; Agam: *Anopheles gambiae*; Hmag: *Hydra magnipapillata*; Nvec: *Nematostella vectensis*; Cele: *Caenorhabditis elegans*; Caps: *Capitella sp*; Hrob: *Helobdella robusta*; Drer: *Danio rerio*; Tnig: *Tetraodon nigroviridis*; Hsap: *Homo sapiens*; Mmus: *Mus musculus*; Lgig: *Lottia gigantea*; Bflo: *Branchiostoma floridae*; Cint: *Ciona intestinalis*; Spur: *Strongylocentrotus purpuratus*; Bmal: *Brugia malayi*; Tadh: *Trichoplax adhaerens*; Sman: *Schistosoma mansoni*; Mbre: *Monosiga brevicollis*; Ehis: *Entamoeba histolytica*; Edis: *Entamoeba dispar*; Einv: *Entamoeba invadens*; Ddis: *Dictyostelium discoideum*; Ppal: *Polysphondylium pallidum*; Dpur: *Dictyostelium purpureum*; Crei: *Chlamydomonas reinhardtii*; Mpus: *Micromonas pusilla*; Micsp: *Micromonas sp*; Chlor: *Chlorella NC64A*; Vcar: *Volvox carteri*; Otau: *Ostreococcus tauri*; Atha: *Arabidopsis thaliana*; Ppat: *Physcomitrella patens*; Cmer: *Cyanidioschyzon merolae*; Smoe: *Selaginella moellendorffii*; Ptri: *Phaeodactylum tricornutum*; Psoj: *Phytophthora sojae*; Pram: *Phytophthora ramorum*; Tpse: *Thalassiosira pseudonana*; Aano: *Aureococcus anophagefferens*; Tthe: *Tetrahymena thermophila*; Ptet: *Paramecium tetraurelia*; Tgon: *Toxoplasma gondii*; Tpar: *Theileria parva*; Tann: *Theileria annulata*; Cpar: *Cryptosporidium parvum*; Cmur: *Cryptosporidium muris*; Pfal: *Plasmodium falciparum*; Bbov: *Babesia bovis*; Pmar: *Perkinsus marinus*; Tcru: *Trypanosoma cruzi*; Tbru: *Trypanosoma brucei*; Lmaj: *Leishmania major*; Linf: *Leishmania infantum*; Ngru: *Naegleria gruberi*; Glam: *Giardia lamblia*; Tvag: *Trichomonas vaginalis*; Gthe: *Guillardia theta*; Ehux: *Emiliania huxleyi*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

UNCORRECTED PROOF

ARTICLE IN PRESS



Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

3 region and the conformational changes arising from GTP sensing, tightens the ring by creating links between the residues forming the circum-nucleotide cage and those closer to the effector interaction interface (Fig. 1, supplementary material). Hence, new features that emerged in Gαs relative to the Arf-like GTPases were primarily related to: 1) interaction with their specific partners such as RGS and Gβ and 2) formation of a structural network to transmit the conformational changes pertaining to the state of the nucleotide (i.e. GTP or GDP) to regions of effector contact. Identification of these residues in Gα also suggests that one of the functions of the RGS domain might be related to increasing the proximity of the separate parts of the ring.

Recognition of these features allowed us to identify Gαs from various eukaryotic lineages that show deviations relative to the canonical conservation pattern (Supplementary material). An example of such a deviant Gα is the plant XLG. The XLG proteins lack the conserved lysine in the Walker A (G1) motif, the arginine finger and the Walker B (G3) aspartate indicating that they are incapable of catalyzing GTP hydrolysis. However, they do possess an otherwise largely intact P-loop and G4 motif indicating that they could still bind a guanine nucleotide, albeit with reduced affinity. Further, they lack the N-terminal helix with the lipid modified glycine, but are fused to a divergent FYVE Zn-finger domain at the N-terminus, which could direct them to sub-cellular localization distinct from the conventional Gαs, such as the nucleus (Supplementary material). We found comparable examples in *Dictyostelium*, where 4 of its 11 Gα paralogs (gpaK, gpaJ, spnA and gbqA) show notable deviations in their conserved motifs. All of them appear to have lost their arginine finger, whereas three of these (gpaJ, spnA and gbqA) have lost either the conserved lysine in the Walker A motif, the aspartate in Walker B, or both. Thus, they are likely to be catalytically inactive (gpaJ, spnA and gbqA) or at best have low activity (gpaK). In spnA the inactive Gα domain is fused to a C-terminal PP2C phosphatase domain and mutational analysis has indicated the importance of this Gα domain for the function of the protein (Aubry and Firtel, 1998). This suggests that in this case the Gα might have been recruited to function as a potential GTP sensor. In gbqA even the G4 motif, which recognizes the guanine of the nucleotide, is considerably divergent, suggesting that it might not bind GTP. We detected a previously unknown Gγ domain N-terminal to the inactive Gα domain in gbqA (Fig. 4B), which suggests that it might function as a regulator of heterotrimeric G-protein signaling by potentially sequestering Gβ subunits (see below). Further, gpaJ, spnA and gbqA have lost the N-terminal helix with the lipid-modified glycine, suggesting that they are unlikely to associate with the membrane. Within the LSE of Gαs in *Naegleria* at least 9 of the paralogs show deviations strongly suggestive of loss of catalytic activity—they have either lost the Walker B aspartate or the Walker A lysine (Supplementary material). At least four show replacement of the arginine finger by a lysine, while four others appear to have entirely lost the arginine finger suggesting that they might have reduced catalytic capability. Thus, at least 38% of Gαs in the *Naegleria* LSE appear to be either catalytic inactive or compromised in activity. This raises the possibility that in *Naegleria* G protein signaling is extensively regulated through the dominant negative regulatory action of these proteins. Unlike the versions in *Dictyostelium,* none of them appear to show fusions to other domains, and most of them show an N-terminal helix with the glycine or cysteines that could target them to the membrane through lipid-modification. Like in certain animal Gαs, representatives from the LSEs observed in mucors, *Naegleria* and *Trichomonas* have distinct N-terminal cysteines that could function as potential sites for alternative S-linked palmitoylation (Escriba et al., 2007). The Gα LSEs in *Caenorhabditis elegans*, mushrooms and *Trichomonas* also show 1–3 versions lacking arginine fingers, which could have lowered activity. Mutational studies in *C. elegans* suggest at least one of them, Gpa-13, might retain a low level of activity and function downstream of the 7TM chemoreceptors (Bastiani and Mendel, 2006) but the details of its biochemistry remain unknown. Our observations suggest that a more careful biochemical analysis of these catalytically compromised or inactive versions might reveal new features of G-protein signaling that were not apparent in the classical models.

3.4. The provenance, structural features and phyletic patterns of the Gβ and Gγ subunits

The Gβ subunit is distinguished from all other WD40-type seven-bladed β-propellers in possessing a distinctive N-terminal bi-helical extension (Lambright et al., 1996). Further, the β-propeller of Gβ shows a circular permutation such that the N-terminal-most strand hydrogen bonds with the three C-terminal-most strands to constitute the 7th blade of the β-propeller (Fig. 2). Gγ has a simple structure, comprised of two helices that interact with the two cognate helices in the bi-helical N-terminus of the Gβ subunit. Interestingly, these bi-helical modules from the Gβ and Gγ subunit structurally align very well with each other (Fig. 1, Supplementary material) suggesting that they are likely to have been derived from a common ancestor. The Gβ subunit shows a characteristic pattern of sequence conservation in its N-terminal bi-helical module and the two C-terminal propeller blades (6 and 7) that are related to its interaction with the Gγ subunit. Similarly, it also shows a specific pattern of sequence conservation in the 1st propeller blade that interacts with the N-terminal helix of the Gα subunit and in all the blades except the 6th blade that interacts with the region just downstream of the switch-2 region of Gα (Figs. 1, 2). These conservation patterns help in precisely identifying the Gβ subunits through sequence analysis and distinguishing them from various WD40 proteins in eukaryotic proteomes that have been mis-annotated as Gβs. While most of the conserved features of Gγ relate primarily to its interactions with Gβ, it possesses an additional C-terminal extended region, which in all standalone versions of the Gγ domain contains one or a pair of cysteines that are farnesylated (Fig. 2) (Escriba et al., 2007; Wensel, 2008). These characteristic features help in identifying Gγ subunits throughout eukaryotes despite their small size and high sequence divergence.

Based on the above conservation patterns we detected Gβ subunits in all eukaryotes with Gαs except in *Trichomonas* (Fig. 3). Interestingly, none of the genomes in which we observed an LSE of Gα, showed a concomitant expansion of the Gβ subunit (Fig. 3). Most genomes encoded only a single paralog of Gβ, while 4-5 paralogs were detectable in *Naegleria*, the mucor *Phycomyces* and vertebrates. This suggests that the diversification of core heterotrimeric GTpase system in eukaryotes occurred primarily via LSEs of Gαs, which are likely to share a much smaller set of Gβ subunits between themselves. Since the Gγ subunits are small and divergent, we initiated Jackhmmer searches with several distinct starting points to identify previously undetected Gγ subunits. Consequently, we recovered previously undetected version in plants, slimemolds and stramenopiles and haptophytes (e.g. *Emiliania* and *Phytophthora;* Supplementary Material). Currently all organisms possessing a Gβ also possess a Gγ, indicating the obligate functional interaction between the two across eukaryotes.

Fig. 4. (A) Domain network graph of RGS-containing proteins. The network is an ordered graph representing the connection between RGS and the domains it is fused to in various polypeptides. The direction of the arrow denotes the relative positions of the domain in the polypeptide from the N-terminus to the C-terminus. The networks are shown for the different clades. (B) Domain architectures of RGS and Gα containing proteins are shown. The 7TM domain is shown as 7 connected TMs. The gene, organism name and gi are shown for proteins with gene names, while only the gi and the organism name are shown for draft genomes with non-standardized gene names. The domain abbreviations are TM—Transmembrane helix; B-Propeller—β-propeller; ACYC—Adenylyl Cyclase; ANK—Ankyrin repeats; ACB—Axin beta-catenin binding domain; BB—B-box zinc finger; B-Propeller—β-propeller; BLUF—Sensors of blue-light using FAD; BLec—BULBLECTIN; CS—Cysteine String; EF—EFHAND; FB—FBOX; Gg—Gγ; GLC—GoLoco; IG—Immunoglobulin domain; Ngru_X—Extra cellular domain specific to *Naegleria* 7TM + RGS proteins; RA—Ras association (RalGDS/AF-6) domain; RBD—Raf-like Ras-binding domain; RCC1—Regulator of chromosome condensation (RCC1) repeat, a variety of β-propeller; SH—S-helix; SIG—Signal peptide; STYKIN—S/T/Y kinase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

ARTICLE IN PRESS

Interestingly, while there are four Gβs in *Naegleria*, no Gγs were detected in the published proteome (Fritz-Laylin et al.). However, translating searches with TBLASTN recovered two standalone Gγ paralogs that possessed all the features typical of eukaryotic Gγs, including those required for interactions with the Gβs (Fig. 2, Supplementary material). Likewise, given that the *Naegleria* Gβs possess the features that are required for Gγ and Gα interaction, the conventional heterotrimeric G-proteins might inferred for this organism. As in the case of Gβ there are no concomitant expansions of Gγs relative to the Gα expansions in eukaryotes. The only exceptions are the vertebrates which have a LSE of Gγs, with at least 13 solo and 6 fused to other domains. Outside of animals the only example of Gγs fused to other domains are those fused to the inactive Gα in slimemolds (Fig. 4). Interestingly, *Trichomonas* was the only organism in which both Gβ and Gγ were not detectable, even via translating searches in the published genome sequence. This is particularly puzzling because the Gα subunits still retain interaction surfaces for Gβ that are seen in the other eukaryotes. Given the potentially early-branching position of *Trichomonas* one possibility is that the conventional Gβ seen in other eukaryotes had not yet emerged. However, in absence of data from other related organisms it is not impossible that these subunits were secondarily lost. Nevertheless, it suggests that *Trichomonas* might possess a unique form of G-protein signal independent of Gβγ subunits.

4. Structural features and phyletic patterns of RGS domain proteins

The RGS, like most other GAP domains of the extended Ras-like clade, appears to be an α-helical domain innovated in the eukaryotes. Its core contains a 4-helical bundle linked to an N-terminal trihelical extension (hence the RGS helices are numbered α1-7, Figs. 1, 2). The 4-helical bundle does not show close structure or sequence relationships to any of the other 4-helical bundles. The C-terminal helix (α7) is the longest and shows a characteristic bend in the helical axis and pronounced distortions of the C-terminal end. A wealth of biochemical and structural data has accumulated for RGS domains from vertebrates, which point to a diversity in their binding modes and catalytic capabilities (Roush, 1996; Siderovski et al., 1996; Tesmer et al., 1997; Chen et al., 2005; Johnston et al., 2008; Soundararajan et al., 2008; Hurst and Hooks, 2009; Kimple et al., 2009; McCoy and Hepler, 2009; Porter and Koelle, 2009). The majority of vertebrate RGS domains are catalytically active as GAPs and co-crystal structures with their cognate Gαs reveals that the common denominator is their interaction with the switch-1 and -2 regions of the latter proteins (Tesmer et al., 1997; Soundararajan et al., 2008). However, different catalytically active RGS domains differ in their degree of engagement of the α-helical insert (after strand 5) and the switch-3 region of Gα. For example, vertebrate RGS4 shows very limited interaction with the α-helical insert, whereas RGS8 and RGS9 show more extensive interactions, albeit mediated by a distinct set of residues (Tesmer et al., 1997; Soundararajan et al., 2008). Similarly, RGS4 and related RGS domains show extensive switch-3 interaction, whereas RGS10 apparently shows a lower degree of interaction (Soundararajan et al., 2008). Beyond these catalytically active versions, several "unconventional" versions have been recognized in animals. Of these, the versions fused to the RhoGEF domains do not conserve the typical interface of the active RGS domains present to the Gαs. Instead they appear to have acquired a neomorphic N-terminal extension that is not conserved in any other RGS proteins by which they interact and elicit GAP activity on Gαs (Chen et al., 2005; Chen et al., 2008). The RGS domains found in the GPCR kinases, axins, sorting nexins, and probably RGS22 and RGSL1 proteins also do not appear to have the typical Gα interaction interface and none of them have been shown to function as active GAPs (Spink et al., 2000; Dajani et al., 2003; Tesmer et al., 2005; Singh et al., 2008; Tesmer et al., 2010). Comparison of catalytically active RGS domains with the atypical versions has revealed several differences (Soundararajan et al., 2008). However, the essential functional differences have been difficult to evaluate because most systematic comparisons have concentrated on active RGS domains from animals, all of which tend to show a high degree of similarity.

We used the new information from the RGS domains from across the eukaryotic tree (Fig. 3, supplementary material) to assess the conservation patterns more robustly. An analysis of the column-wise entropy using a comprehensive alignment of all types of RGS domain across eukaryotes revealed a clear pattern with relatively few strongly conserved positions (Fig. 2, Supplementary material). A subset of these low entropy positions are hydrophobic and are central to the packing of the helical bundle. However, others occupy more exposed locations and are typically polar residues that overlap with those positions which make key contacts with the core Gα catalytic domain in crystal structures of active RGS domains (Figs. 1, 2; PDB: 1agr, 2gtp, 2ihb, 2ik8 and 2ode, Supplementary material) (Tesmer et al., 1997; Johnston et al., 2007; Soundararajan et al., 2008). These positions include: 1) A conserved glutamate (E87 in the human RGS4 structure 1agr) in α4 of the RGS domain that interacts with the characteristic lysine in the Gα switch-2 region. The position immediately downstream of this glutamate is also well-conserved in the conventional active RGS domains and is typically an asparagine or serine (N88 in PDB: 1agr) that interacts with the residue immediately upstream of the sensor threonine in the switch-1 of Gα. Thus, these two positions of the RGS bridge the switch-1 and switch-2 regions. 2) A conserved polar position, typically an asparagine (N128 in RGS4, PDB: 1agr) interacts extensively with the switch-2 region of Gα, including the glutamine residue of the DGGQ motif (Fig. 2) and the conserved residues that are respectively two and three positions downstream of this glutamine. This conserved RGS position is also reasonably close to the backbone of the sensor threonine of switch-1 in several structures. 3) The conserved polar residue in α7, usually an aspartate or asparagine (D163 in RGS4, PDB: 1agr) makes a key interaction with the backbone of the sensor threonine of switch-1. 4) The arginine in α7 (R167 in RGS4, PDB: 1agr), which is 4 positions downstream of the previously mentioned residue interacts with that aspartate or asparagine and helps position it suitably for its Gα interaction. Our analysis based on the pan-eukaryotic collection of RGS domains reveals that of all the residues identified in the Gα–RGS interface, those in the above positions represent the minimal set of conserved interactions between the RGS and Gα and were present in the ancestral RGS domain. Given their consistent presence in active RGS domains they also help define the versions of the domain with conventional GAP activity. The remaining positions of the RGS domain, which mediate interactions with the α-helical insert and the switch-3 region, are more variable in the pan-eukaryotic set of RGS domains than those from animals alone (which includes the available crystal structures; Fig. 2 and supplementary material) (Sierra et al., 2002). This suggests that the variability in these interfaces observed in the recent crystal structures is likely to represent only a subset of the actual diversity of these interactions across eukaryotes. Hence, it is likely that the interactions with switch-3 and the α-helical insert are important determinants of the specificity for different target Gαs and the subtle differences in the regulatory properties between RGS domains.

Analysis of the phyletic patterns of the RGS proteins revealed a strict correspondence to the presence of Gα in a given organism (Fig. 3). Furthermore, the number of RGS domains is positively correlated with the number of Gα paralogs across eukaryotes (Fig. 5, $R^2 = 0.77$ for a linear fit). This suggests that the expansion of the Gαs in particular lineage is typically accompanied by a corresponding expansion of the RGS domain proteins. The only exceptions to this trend are the mushrooms (e.g. *Laccaria* and *Coprinopsis*), which show major lineage-specific expansion of Gαs (17–41 paralogs) but not of the RGS domains (only three paralogs). This excess of Gαs arises from the LSEs occurring in the single fungus-specific clade of Gαs that is also expanded in chythrids (Supplementary material). One possibility is that the excess of Gαs in mushrooms signal through RGS-independent signaling pathways. However, they do not show drastic divergence or any particular deviations in the RGS-interacting regions. This suggests that

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

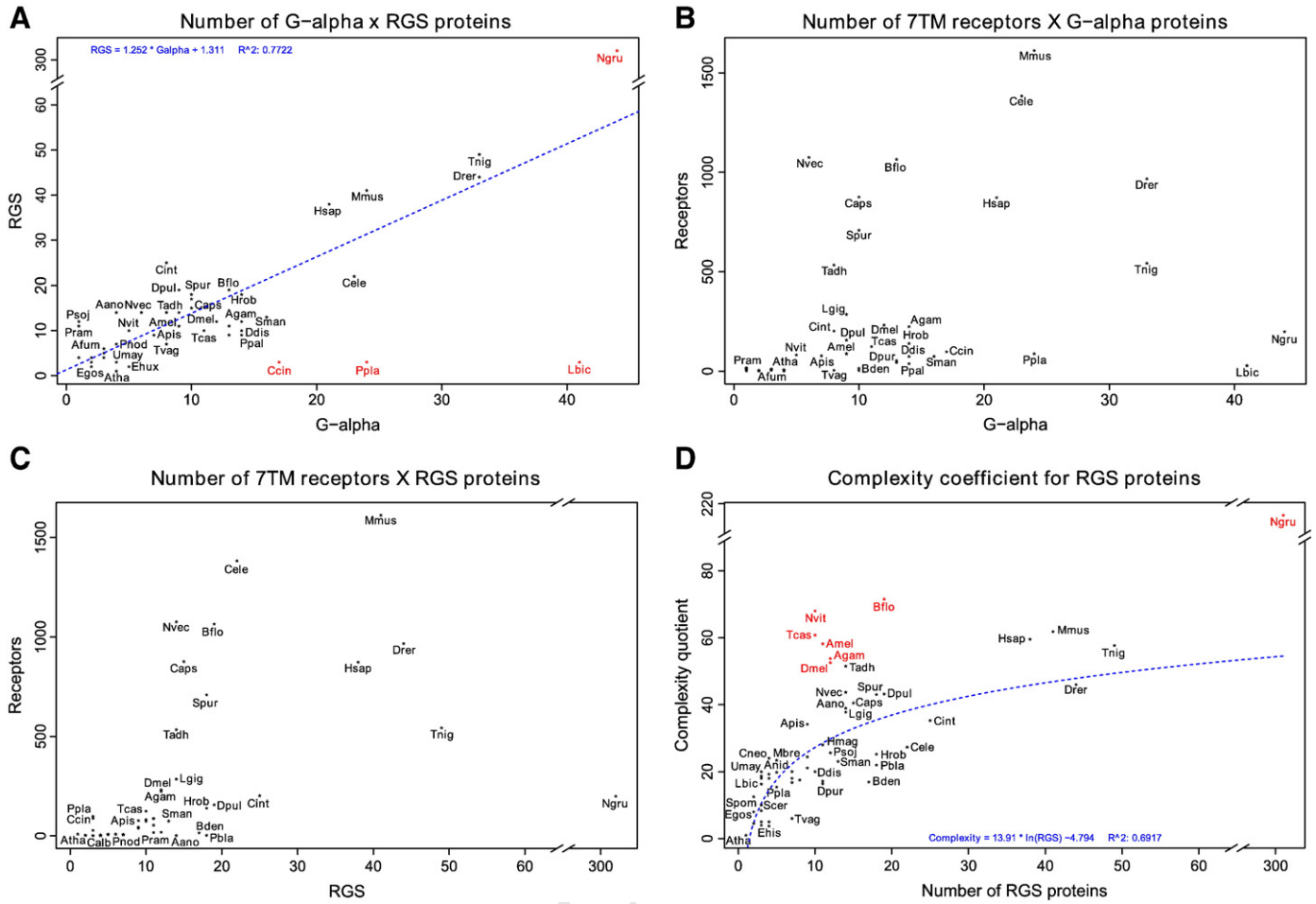


Fig. 5. (A) Linear scaling of total number of RGS proteins against Gα proteins in eukaryotes along with the best-fit line without the outlying organisms (red). (B) Scatter plot of the number of 7TM receptor containing proteins with Gα proteins in eukaryotes. (C) Scatter plot of the number of 7TM receptor containing proteins with RGS proteins in eukaryotes. (D) Complexity quotient plot for RGS proteins. The complexity quotient for an organism is defined as the product of two values: the number of different types of domains that co-occurs in signaling proteins, and the average number of domains detected in these proteins. The complexity quotient is plotted against the total number of RGS-containing proteins in a given organism. A saturation curve fitting the distribution without the outlying organisms is shown in red. Organism abbreviations are as in Fig. 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the expansion of Gα subunits in the mushrooms possibly functions with the same limited set of RGS proteins and the expansion of the Gαs is unlikely to represent an increase in terms of the number of functionally distinct signaling cascades initiated by them. Given the role for Gαs in sex pheromone signaling in fungi (Versele et al., 2001; Koelle, 2006), we speculate that these distinct Gαs have a relationship with the diversity in the mating types observed in these fungi. Thus, the different Gαs are postulated to function as alternative partners with the same limited set of RGS proteins in different mating-type backgrounds. Phylogenetic analysis revealed that the evolutionary history of the RGS domain is dominated by several independent lineage-specific expansions across eukaryotes (Supplementary material, Fig. 3). A notable expansion is inferred to have occurred in the common ancestor of metazoans (Sierra et al., 2002)—conservatively, 4 classical catalytically active versions, 4 sorting nexin-associated RGS domains, 2 GPCR kinase-associated versions, 1 axin-like version, and probably 1 RhoGEF- associated version (a total of 12) are inferred as being present in the common ancestor of all metazoans. In contrast, the common ancestor of fungi is conservatively inferred as containing three RGS domains (two catalytically active and one sorting nexin associated version). The sorting nexin-associated version appears to have diverged from the catalytically active versions prior to the divergence of fungi and animals. Within metazoa, several additional independent expansions are observed in lineages such as vertebrates, nematodes and urochordates, with more than half the RGS domains in these lineages emerging as result of independent expansions. Outside of metazoa major LSEs are observed in chythrid fungi, and basal eukaryotes such as *Trichomonas* and *Naegleria*, with the latter organism containing the largest complement of RGS domains observed in any organism to date (at least 302 paralogous copies, of which at least 85% are predicted to be catalytic active versions; Fig. 3). These observations further strengthen the contention that across eukaryotes the RGS domain function is closely related to the regulation of Gα signaling and that the inactive versions linked to other process (e.g. the sorting nexins) are secondary adaptations of this domain. The pattern of multiple independent LSEs of RGS domains indicates that this represents the primary evolutionary mechanism by which GPCRs, and the stimuli sensed by them, have been recruited to impact a diverse set of intracellular processes across eukaryotes. Thus, the RGS domain proteins are the primary channel for the Gαs with the GPCR-linked sensory inputs being utilized in lineage-specific adaptations of organisms to their external environment and physiology.

5. Architectural diversity of RGS domain proteins

Consistent with the above inference that RGS proteins are central to the lineage-specific adaptations of G-protein mediated signaling, they show the greatest diversity in domain architectures of all the components of the heterotrimeric G-protein signaling system. Hence,

ARTICLE IN PRESS

we systematically investigated their domain architectures to understand more clearly how they have been utilized in their lineage-specific signaling roles. We have previously used the complexity quotient, which takes into account both the number of domains and the number of *different types* of domains in proteins of a given category, as a measure of domain architectural complexity (Fig. 5) (Aravind et al., 2001; Burroughs et al., 2007). The plot of complexity quotient against the number of proteins with the domain, across completely sequenced eukaryotes, indicates that the complexity of architectures of RGS domain proteins increases approximately as a logarithmic function of the number of proteins with the domain. This trend is comparable to similar logarithmic increases previously observed for other "signaling" domains occurring in multi-domain architectures, such as the ubiquitin-like domains (Burroughs et al., 2007). Alternatively, the domain architectures can be represented as an ordered graph which depicts the connectivity between the domains as directed edges (Anantharaman et al., 2007) (Fig. 4). These graphs reinforce the above picture, with lineages with greater number of RGS domains also showing much denser graphs spanning a greater number of connections with other distinct domains. In phyletic terms the most diverse set of architectures are seen in *Naegleria*, followed by metazoa, and then by fungi, stramenopiles and amoebozoans (Fig. 4). This observation, taken together with evidence for independent lineage-specific expansion of the Gα and the RGS domains in metazoans and *Naegleria*, suggests that large-scale propagation of the heterotrimeric G-protein signal through a diverse set of intracellular channels evolved independently at least twice in eukaryotic evolution. In contrast to the above lineages, *Trichomonas* and plants have a very limited architectural diversity in their RGS proteins, suggesting that in these organisms heterotrimeric G-protein signals propagate via a relatively small set of intracellular cascades.

Interestingly, these graphs in combination with sequence analysis suggest that certain associations of the RGS domain with other domains might have arisen independently on more than one occasion. A notable example is the independent fusion of the RGS domain to protein kinase domains in metazoa on one hand and in amoebozoa and stramenopiles on the other. Previously it was suggested that the *Dictyostelium*-type RGS-kinase proteins might be related to the metazoan GPCR kinases (Johnston et al., 2008). However, our analysis indicated that this is not the case, with multiple fusions between the kinase and RGS domains having arisen independently in the amoebozoans. Unlike in the metazoan proteins, the kinase-linked RGS domains in both the amoebozoans and stramenopiles are predicted to be catalytically active (Supplementary material), indicating that they represent a functionally distinct theme from the GPCR kinases (Tesmer et al., 2005; Singh et al., 2008; Tesmer et al., 2010). Architectures of RGS domain proteins recovered from early-branching eukaryotes point to multiple independent associations between the RGS domain and several distinct lipid-binding domains throughout eukaryotic evolution (Fig. 4). In the fungi-metazoa clade fusions at least 7 domains, namely C1, C2, Cysteine string (via palmitoylation) PH, PX, PXA and TM segments, which could mediate interactions with membranes have been observed. Convergent fusions to the PH domain are also observed stramenopiles and *Naegleria,* and to the C2 and PX domains in *Naegleria* (Fig. 4). Additionally, *Naegleria* encodes at least 54 proteins in which the RGS domain is fused to a C-terminal lipid-binding START domain (Ponting and Aravind, 1999; Iyer et al., 2001), a combination which is currently known from no other organism. Previous studies in animals and fungi indicate that the localization of RGS proteins to the inner surface of the cell-membrane results in enhanced interactions with the heterotrimeric G-protein signaling complexes and allow for maximal GAP activity (Srinivasa et al., 1998; Tu et al., 2001; Roy et al., 2003). In light of the above observations on domain architectures, it appears likely that membrane-association has played a major role throughout eukaryotic evolution in selecting for linkages between RGS and lipid-binding domains. In particular, the association with the START domain in *Naegleria*, might have a special role in recruiting heterotrimeric G-protein signaling to specific membrane microdomains enriched in sterol lipids which can be sensed by the START domain (Alpy and Tomasetto, 2005). The RGS protein architectures from early-branching eukaryotes also point to multiple independent origins for linkages between heterotrimeric G-protein signaling and cyclic nucleotide signaling in eukaryotic evolution. In metazoa, Gs- and Gi-type Gαs have evolved features that allow specific interactions with their adenylyl cyclase effectors (Berlot and Bourne, 1992; Sunahara and Taussig, 2002; Mou et al., 2009). However, the Gs- and Gi- type Gαs are a part of the radiation of Gαs that occurred early in animal evolution and do not have cognates which conserve the adenylyl cyclase interaction interface outside of metazoa. Likewise, the vertebrate RGS2, which interacts with both Gs and the adenylyl cyclases to attenuate cAMP production, belongs to a radiation of RGS domains that occurred only in the vertebrates (Roy et al., 2003; Kimple et al., 2009). Thus, the regulation of cyclic nucleotide signaling by Gs, Gi and RGS2 appears to have emerged only within metazoa. However, we observed that in *Naegleria* there are at least two bacterial-type cNMP cyclases, which are fused to C-terminal RGS domains (Fig. 4, Supplementary Material), suggesting a functional interaction with Gα-dependent signaling. It is conceivable that the fused RGS domain functions as an in built regulatory mechanism that helps the cyclase to limit Gα signaling either by GAP activity or by steric hindrance, as has been proposed for RGS2 (Roy et al., 2003). Another independent case of functional linkage of these systems is seen in the stramenopiles, which display a conserved architecture combining a RGS domain to two N-terminal cNMP-binding domains (Fig. 4, Supplementary Material). Hence, these proteins could potentially regulate Gα signaling in a cyclic nucleotide-dependent manner.

Examination of the actual domain architectures of RGS proteins from the previously uncharacterized eukaryotic lineages reveals several novel linkages that greatly expand the diversity of regulatory roles to which heterotrimeric G-protein signaling has been recruited (Fig. 4). The most striking of these are novel bacterial-type receptor proteins from *Naegleria* with intracellular RGS domains. While previously characterized receptors with RGS domain, such as the plant sugar receptors (e.g. *Arabidopsis* RGS1), are conventional 7TM receptors, these *Naegleria* proteins exhibit a different architecture – they possess two TM regions bracketing different types of extracellular small-molecule-binding sensor domains, which are typical of bacterial receptors, followed by an intracellular module containing an RGS domain (Fig. 4). Further, just as in the case of their bacterial cognates, their intracellular module might exhibit bacterial-type signal-transmission domains such as HAMP or PAS. The extracellular sensor domains of these *Naegleria* proteins are of three types: 1) At least 52 of them contain a 4-helical bundle domain typical of bacterial chemotactic receptors that has been called the "CHASE3 domain" (Zhulin et al., 2003; Aravind et al., 2010). 2) At least 10 of them contain a divergent version of a PAS domain previously called "CHASE4" (Zhulin et al., 2003; Aravind et al., 2010). 3) At least two contain a CHASE domain which also possesses the same fold as the PAS domains (Anantharaman and Aravind, 2001; Mougel and Zhulin, 2001). The above features strongly suggest that the progenitors of these receptors where acquired by heterolobosea (the clade containing *Naegleria)* from bacteria via lateral transfer. Their combinations with RGS domains appear to have converted them into a novel type of receptors that recruit Gαs. Hence, they represent a remarkable example of a receptor of bacterial origin being recruited to signal via a mechanism of eukaryotic provenance by the fusion of a single domain. Also notable is the fusion in *Naegleria* of the RGS domain to the flavin-binding BLUF (Gomelsky and Klug, 2002) (two proteins) and the thioredoxin-like (11 proteins) domains (Fig. 4). These latter domains could potentially function as sensors of changes in redox potential and might represent a hitherto unknown mechanism of signaling such changes via the heterotrimeric G-protein system. In photosynthetic stramenopiles such as *Aureococcus* and *Ectocarpus* we found fusions of RGS domain to EF-hand domains (Fig. 4), suggesting

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

ARTICLE IN PRESS

that in these organisms there could be unique cross-links between calcium-dependent and G-protein signaling.

6. 7TM receptors with and without C-terminal RGS domains and the evolution of functional linkages between 7TM receptors and heterotrimeric G-protein signaling

Examination of the architectures suggested a striking qualitative difference between animals and many of other eukaryotic lineages: clades such as plants, fungi, *Naegleria* and *Trichomonas* contain one to several representatives of 7TM receptors with C-terminal intracellular RGS domains (Fig. 4) (Johnston et al., 2007). In these clades all or a notable fraction of the RGS domains occur fused to the 7TM receptor (Fig. 3). In plants, fungi and *Trichomonas* these proteins contain just a simple 7TM receptor domain and a C-terminal RGS domain (Johnston et al., 2007; Grigston et al., 2008; Johnston et al., 2008). In *Naegleria* over one third (at least 115 proteins) of the RGS domains are linked to the 7TM domain and there are several more complex receptor architectures that are characterized by the fusion of further extracellular modules to the N-terminus of the 7TM domain. These include domains involved in adhesive interactions like the EGF, polycystic kidney disease (PKD) and bulb-lectin domains, the small-molecule binding PBP-II domain and a previously uncharacterized *Naegleria*-specific domain (Fig. 4, Supplementary material). The combination of a 7TM domain with the PBP-II domain is reminiscent of a similar combination found in the metabotropic glutamate and taste receptors of animals (Pierce et al., 2002) while those fused to the adhesion-related domains resemble the adhesion-GPCRs of animals (Yona et al., 2008). However, these *Naegleria* proteins do not display a close relationship to these latter animal receptors suggesting that this combination emerged independently at least twice during the evolution of GPCRs. The diverse architectures suggest that *Naegleria* might utilize these 7TM both to recognize solutes (via the PBP-II domain) as well as surfaces, such as those of prey and substrata (via the adhesion modules). The sporadic distribution of these 7TM-linked RGS proteins in eukaryotes points to a history involving gene losses and also potentially lateral transfers. Indeed, the plant proteins strongly group with the *Trichomonas* versions to the exclusion of all other RGS + 7TM proteins in phylogenetic trees, supporting the possibility of lateral transfer, although the direction is currently unclear (Supplementary material). Nevertheless, this architecture appears to have independently expanded in *Naegleria*, *Trichomonas* and chythrid fungi, making it the dominant mode of G-protein signaling in these organisms. In stark contrast to these clades, despite possessing a large number of both RGS domains and 7TM receptors there is not even a single instance of a fusion between a RGS domain and a 7TM receptor in the animal lineage. Like animals, all the lineages with 7TM + RGS proteins also contain conventional 7TM receptors without RGS domains (Jones, 2002; Grigston et al., 2008; Johnston et al., 2008) (Fig. 3). At least in the case of the fungi there is evidence that these 7TM receptors without RGS domains function as GEFs, as in the case of the animal receptors (Siderovski and Willard, 2005; Koelle, 2006). Thus, it appears that in most eukaryotes there are different 7TM receptors, which can function either as GEFs or as GAPs, whereas in animals there has been a complete shift towards GEF-type receptors. Further, with the exception of plants, we observed that all these other lineages also encode standalone RGS domain proteins. Studies in plants have hinted that the rate limiting step for their G-proteins is the GTP hydrolysis step, rather than exchange of GTP for GDP, which is the case in animals (Johnston et al., 2007; Grigston et al., 2008). In light of the above observations it is possible that most eukaryotic lineages might possess Gαs that might have either the GDP exchange or GTP hydrolysis as a limiting step and accordingly function better with either a GAP- or GEF-type 7TM receptor. It appears that prior to the lineage-specific expansion of Gαs in the animal lineage their ancestral version became one with a relatively low rate constant for GDP exchange. The propagation of this property across all descendent animal Gαs appears to have eventually favored only GEF-type 7TM receptors in this lineage. A similar situation appears to have independently prevailed in the amoebozoans and stramenopiles.

The above picture indicates that the role of 7TM receptors and the nature of their interactions with heterotrimeric G-proteins appear to have undergone notable shifts within particular lineages and prompted us to more closely examine the evolution of this functional linkage. Prokaryotic sensory rhodopsins and our earlier discovery of several distinct types of bacterial 7TM signaling receptors suggested that the origins of the 7TM receptors ultimately lay in the bacterial world (Anantharaman and Aravind, 2003; Jekely, 2009). The bacterial 7TM-receptors (other than the sensory rhodopsins) architecturally resemble the *Naegleria* 7TM + RGS proteins in that they typically have an extracellular sensor domain, followed by a 7TM domain linked to a C-terminal intracellular signaling domain. However, these signaling domains are unrelated to G-protein signaling; instead they span a diversity of bacterial-type signaling domains such histidine kinases, phosphatases, HD-hydrolases and domains involved in diguanylate production and degradation (Anantharaman and Aravind, 2003). Even within eukaryotes, in various lineages such as animals and slime molds, there is evidence for 7TM receptors signaling independently of the G-proteins via pathways such as MAP kinase and glycogen synthetase kinase-3 (Brzostowski and Kimmel, 2001; Pierce et al., 2002). To investigate the degree of the functional linkage between G-protein signaling and 7TM receptors in eukaryotes we compared the phyletic profiles of these families. The number of 7TM receptors encoded by an organism is generally positively correlated with the numbers of both Gαs and RGS domain proteins (Fig. 5). Thus, various evolutionarily distinct lineages, such as animals, chythrid fungi and *Naegleria*, with large complements of RGS domain proteins also have notably expanded complements of 7TM receptors (Fig. 3). Furthermore, every organism encoding heterotrimeric GTPase subunits and RGS domains necessarily encode 7TM receptors. These observations point to a strong link between the two throughout eukaryotic evolution. The presence of both 7TM receptors and a part of the heterotrimeric G-protein system in *Trichomonas* indicates that, while the 7TM-dependent signaling originally arose independently of G-proteins, it was probably coupled to the Gα and RGS proteins even before the last eukaryotic common ancestor (LECA). The exact nature of the 7TM landscape in the LECA is much harder to reconstruct on account of most of 7TM receptor evolution in eukaryotes being dominated by lineage-specific expansion (Rompler et al., 2007). However, given the presence of both RGS-fused and standalone versions of 7TM receptors in a wide range of eukaryotes, i.e., plants, fungi, heteroloboseans (*Naegleria*) and parabasalids (*Trichomonas*), it is possible that both types were already present in the LECA. The RGS fused versions show a distant relationship with the metabotropic glutamate/taste receptor-like proteins (the so called "family C") (Pierce et al., 2002) suggesting that the later type could have emerged secondarily from a precursor of the former type. The majority of receptors in the clade including animals, fungi and plants belong to the group typified by the visual rhodopsins and the animal chemoreceptors (the so called "family A") (Pierce et al., 2002). The evidence from *Naegleria* and *Trichomonas* suggests that these were perhaps either absent or in low numbers in LECA and expanded only later in eukaryotic evolution. While certain other eukaryotic families with 7 TM segments, such as the LUSTR (GPR107/GPR180) family (Edgar, 2007) are traceable to LECA, there is no evidence in support of them functioning as receptors. More recently a family of *Arabidopsis* TM proteins (GTG1/2) were claimed to be GPCRs for abscissic acid with intrinsic GTPase activity (Pandey et al., 2009). However, sequence analysis reveals that the claim for GTPase activity was based on a spurious alignment and statistically unsupported inference of relationship to G-proteins. Instead, independent evidence suggests

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

ARTICLE IN PRESS

that this conserved family of membrane proteins is likely to function as a voltage-gated ion-channel that promotes acidification of the Golgi compartment (Maeda et al., 2008). Hence, it is highly unlikely that these proteins have any role as receptors in GTP-dependent signaling.

We uncovered at least one eukaryotic lineage, namely the alveolates, which contains 7TM receptors but entirely lacks heterotrimeric GTPase signaling (Fig. 3). In alveolates, 7TM proteins related to the classical eukaryotic GPCRs (i.e. closest to the so called "family A") are observed in ciliates, *Perkinsus* and most of the apicomplexans including the malarial parasite *Plasmodium*, with modest lineage-specific expansions in the ciliates (Supplementary material). Their closest sister group, the chromists, have both TM receptors and GPCRs. This pattern suggests that while the 7TM receptors were retained and in some cases even expanded, the heterotrimeric G-protein system was entirely lost, probably in the common ancestor of the alveolates themselves. This exception to the general pattern of coupled presence of the two signaling elements indicates that the eukaryotes have probably retained a G-protein-independent signaling mechanism via 7TM receptors throughout their evolution. It is likely these independent mechanisms have taken the dominant role in the alveolates after the loss of the heterotrimeric G-proteins themselves.

7. General considerations and conclusions

A key realization of the post-genomic era has been that several major signaling systems that were considered quintessentially eukaryotic are actually of ultimately bacterial provenance (Aravind et al., 2003; Iyer et al., 2006; Aravind et al., 2009). In face of this it is interesting to note that the wealth of genomic data reinforces a eukaryotic provenance for heterotrimeric GTPase signaling, although the 7TM receptors themselves might have first emerged in bacteria. The analysis presented here suggests that the Gα subunit of the heterotrimeric G-proteins emerged prior to the radiation of the extant eukaryotes from an ARF-like GTPase, which was in turn derived from the MglA-like precursor from one of the prokaryotic progenitors of the ancestral eukaryote. Further, this analysis also suggests that the helical N-terminal extension of Gβ and Gγ subunit likely share a common origin. This raises the question as to how the Arf-like proteins, which were clearly involved in trafficking of molecules by membrane vesicles (Li et al., 2004; Kahn et al., 2006; Kahn, 2009) were recruited for receptor-linked signaling to give rise to the ancestral Gα. The role of the ARF-like GTPases in the trafficking of membrane proteins through vesicles would have provided them the initial opportunity to interact with their cargo, which might have included early 7TM receptors. Further, it should be noted that in the formation of vesicles that transport cargo back and forth from the ER and Golgi, the Arf-like proteins interact with coatomer cage proteins which contain WD40 repeats (Lee and Goldberg, 2010). Hence, it is conceivable that the Gβ subunit emerged from such a β-propeller precursor.

The results presented here suggest that inferences regarding heterotrimeric GTPase signaling from the model animals, fungi and plants are only the tip of the iceberg with respect to the understanding of this system. The data from the newly sequenced eukaryotic genomes helps in greatly refining the set of residues involved in key functional interactions amongst the components of this system. These predictions lend themselves to future experimental tests which could help in improving our understanding of G-protein function even in the model systems. Further, the wealth of new connections uncovered in the basal eukaryotes suggests that organisms such as *Naegleria* and chythrids might offer new vistas to experimentally explore G-protein function. In particular, these systems appear to contain a predominance of receptors that function as GAPs rather than GEFs as in the case of animals. Hence, they could help in better understanding how the catalytic properties of the Gαs might select for alternative receptors types. Furthermore, the novel types of receptors identified in *Naegleria* might open new windows into understanding how G-protein signaling might work in non-7TM contexts. Finally, *Trichomonas* could also act as a potential model to study G-protein signaling independent of Gβγ, whereas alveolates might provide models for G-protein-independent 7TM signaling in eukaryotes.

Supplementary materials related to this article can be found online at doi:10.1016/j.gene.2010.12.001.

Acknowledgments

The work by the authors was supported by the intramural funds of the National Library of Medicine, NIH. We would like to thank Lakshminarayan M. Iyer for his role in an early stage of this early work and for reading the final manuscript.

References

Abramowitz, J., Iyengar, R., Birnbaumer, L., 1979. Guanyl nucleotide regulation of hormonally-responsive adenylyl cyclases. Mol. Cell. Endocrinol. 16, 129–146.

Alpy, F., Tomasetto, C., 2005. Give lipids a START: the StAR-related lipid transfer (START) domain in mammals. J. Cell Sci. 118, 2791–2801.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Anantharaman, V., Aravind, L., 2001. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem. Sci. 26, 579–582.

Anantharaman, V., Aravind, L., 2003. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. BMC Genomics 4, 34.

Anantharaman, V., Iyer, L.M., Aravind, L., 2007. Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation. Annu. Rev. Microbiol. 61, 453–475.

Aravind, L., Anantharaman, V., Iyer, L.M., 2003. Evolutionary connections between bacterial and eukaryotic signaling systems: a genomic perspective. Curr. Opin. Microbiol. 6, 490–497.

Aravind, L., Anantharaman, V., Venancio, T.M., 2009. Apprehending multicellularity: regulatory networks, genomics, and evolution. Birth Defects Res. C Embryo Today 87, 143–164.

Aravind, L., Dixit, V.M., Koonin, E.V., 2001. Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. Science 291, 1279–1284.

Aravind, L., Iyer, L.M. and Anantharaman, V.: Natural history of sensor domains in bacterial signaling systems In: Spiro, S. and Dixon, R. (Spiro, S. and Dixon, R. (Spiro, S. and Dixon, R.s), Sensory Mechanisms in Bacteria: Molecular Aspects of Signal Recognition. Caister Academic Press, London (2010), pp. 280.

Arisue, N., Hasegawa, M., Hashimoto, T., 2005. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. Mol. Biol. Evol. 22, 409–420.

Aubry, L., Firtel, R.A., 1998. Spalten, a protein containing Galpha-protein-like and PP2C domains, is essential for cell-type differentiation in Dictyostelium. Genes Dev. 12, 1525–1538.

Bastiani, C., Mendel, J., 2006. Heterotrimeric G proteins in *C. elegans*. *WormBook*, pp. 1–25.

Berlot, C.H., Bourne, H.R., 1992. Identification of effector-activating residues of Gs alpha. Cell 68, 911–922.

Berman, D.M., Wilkie, T.M., Gilman, A.G., 1996. GAIP and RGS4 are GTPase-activating proteins for the Gi subfamily of G protein alpha subunits. Cell 86, 445–452.

Bourne, H.R., Sanders, D.A., McCormick, F., 1991. The GTPase superfamily: conserved structure and molecular mechanism. Nature 349, 117–127.

Brzostowski, J.A., Kimmel, A.R., 2001. Signaling at zero G: G-protein-independent functions for 7-TM receptors. Trends Biochem. Sci. 26, 291–297.

Burroughs, A.M., Balaji, S., Iyer, L.M., Aravind, L., 2007. Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. Biol. Direct 2, 18.

Cassel, D., Selinger, Z., 1976. Catecholamine-stimulated GTPase activity in turkey erythrocyte membranes. Biochim. Biophys. Acta 452, 538–551.

Chen, J.G., Willard, F.S., Huang, J., Liang, J., Chasse, S.A., Jones, A.M., Siderovski, D.P., 2003. A seven-transmembrane RGS protein that modulates plant cell proliferation. Science 301, 1728–1731.

Chen, Z., Singer, W.D., Danesh, S.M., Sternweis, P.C., Sprang, S.R., 2008. Recognition of the activated states of Galpha13 by the rgRGS domain of PDZRhoGEF. Structure 16, 1532–1543.

Chen, Z., Singer, W.D., Sternweis, P.C., Sprang, S.R., 2005. Structure of the p115RhoGEF rgRGS domain-Galpha13/i1 chimera complex suggests convergent evolution of a GTPase activator. Nat. Struct. Mol. Biol. 12, 191–197.

Cho, H., Kehrl, J.H., 2009. Chapter 9 regulation of immune function by G protein-coupled receptors, trimeric G proteins, and RGS proteins. Prog. Mol. Biol. Transl. Sci. 86, 249–298.

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T., Bader, G.D., 2007. Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. 2, 2366–2382.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

Cuff, J.A., Barton, G.J., 2000. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins 40, 502–511.

Dajani, R., Fraser, E., Roe, S.M., Yeo, M., Good, V.M., Thompson, V., Dale, T.C., Pearl, L.H., 2003. Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex. EMBO J. 22, 494–501.

De Vries, L., Mousli, M., Wurmser, A., Farquhar, M.G., 1995. GAIP, a protein that specifically interacts with the trimeric G protein G alpha i3, is a member of a protein family with a highly conserved core domain. Proc. Natl Acad. Sci. USA 92, 11916–11920.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

Dong, J.H., Wen, J.F., Tian, H.F., 2007. Homologs of eukaryotic Ras superfamily proteins in prokaryotes and their novel phylogenetic correlation with their eukaryotic analogs. Gene 396, 116–124.

Druey, K.M., Blumer, K.J., Kang, V.H., Kehrl, J.H., 1996. Inhibition of G-protein-mediated MAP kinase activation by a new mammalian gene family. Nature 379, 742–746.

Edgar, A.J., 2007. Human GPR107 and murine Gpr108 are members of the LUSTR family of proteins found in both plants and animals, having similar topology to G-protein coupled receptors. DNA Seq. 18, 235–241.

Elzie, C.A., Colby, J., Sammons, M.A., Janetopoulos, C., 2009. Dynamic localization of G proteins in *Dictyostelium discoideum*. J. Cell Sci. 122, 2597–2603.

Escriba, P.V., Wedegaertner, P.B., Goni, F.M., Vogler, O., 2007. Lipid–protein interactions in GPCR-associated signaling. Biochim. Biophys. Acta 1768, 836–852.

Freissmuth, M., Casey, P.J., Gilman, A.G., 1989. G proteins control diverse pathways of transmembrane signaling. FASEB J. 3, 2125–2131.

Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J.B., Carpenter, M.L., Field, M.C., Kuo, A., Paredez, A., Chapman, J., Pham, J., Shu, S., Neupane, R., Cipriano, M., Mancuso, J., Tu, H., Salamov, A., Lindquist, E., Shapiro, H., Lucas, S., Grigoriev, I.V., Cande, W.Z., Fulton, C., Rokhsar, D.S. and Dawson, S.C. The genome of Naegleria gruberi illuminates early eukaryotic versatility. *Cell* **140**, pp. 631-42. Q2

Gomelsky, M., Klug, G., 2002. BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. Trends Biochem. Sci. 27, 497–500.

Grigston, J.C., Osuna, D., Scheible, W.R., Liu, C., Stitt, M., Jones, A.M., 2008. D-Glucose sensing by a plasma membrane regulator of G signaling protein, AtRGS1. FEBS Lett. 582, 3577–3584.

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

Hanson, M.A., Stevens, R.C., 2009. Discovery of new GPCR biology: one receptor structure at a time. Structure 17, 8–14.

Holm, L., Sander, C., 1998. Touring protein fold space with Dali/FSSP. Nucleic Acids Res. 26, 316–319.

Hurst, J.H., Hooks, S.B., 2009. Regulator of G-protein signaling (RGS) proteins in cancer biology. Biochem. Pharmacol. 78, 1289–1297.

Iyer, L.M., Anantharaman, V., Wolf, M.Y., Aravind, L., 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int. J. Parasitol. 38, 1–31.

Iyer, L.M., Burroughs, A.M., Aravind, L., 2006. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. Genome Biol. 7, R60.

Iyer, L.M., Koonin, E.V., Aravind, L., 2001. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43, 134–144.

Janetopoulos, C., Jin, T., Devreotes, P., 2001. Receptor-mediated activation of heterotrimeric G-proteins in living cells. Science 291, 2408–2411.

Jekely, G., 2003. Small GTPases and the evolution of the eukaryotic cell. Bioessays 25, 1129–1138.

Jekely, G., 2009. Evolution of phototaxis. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 2795–2808.

Johnston, C.A., Taylor, J.P., Gao, Y., Kimple, A.J., Grigston, J.C., Chen, J.G., Siderovski, D.P., Jones, A.M., Willard, F.S., 2007. GTPase acceleration as the rate-limiting step in Arabidopsis G protein-coupled sugar signaling. Proc. Natl Acad. Sci. USA 104, 17317–17322.

Johnston, C.A., Willard, M.D., Kimple, A.J., Siderovski, D.P., Willard, F.S., 2008. A sweet cycle for Arabidopsis G-proteins: recent discoveries and controversies in plant G-protein signal transduction. Plant Signal. Behav. 3, 1067–1076.

Jones, A.M., 2002. G-protein-coupled signaling in Arabidopsis. Curr. Opin. Plant Biol. 5, 402–407.

Kahn, R.A., 2009. Toward a model for Arf GTPases as regulators of traffic at the Golgi. FEBS Lett. 583, 3872–3879.

Kahn, R.A., Cherfils, J., Elias, M., Lovering, R.C., Munro, S., Schurmann, A., 2006. Nomenclature for the human Arf family of GTP-binding proteins: ARF, ARL, and SAR proteins. J. Cell Biol. 172, 645–650.

Kall, L., Krogh, A., Sonnhammer, E.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.

Kimple, A.J., Soundararajan, M., Hutsell, S.Q., Roos, A.K., Urban, D.J., Setola, V., Temple, B. R., Roth, B.L., Knapp, S., Willard, F.S., Siderovski, D.P., 2009. Structural determinants of G-protein alpha subunit selectivity by regulator of G-protein signaling 2 (RGS2). J. Biol. Chem. 284, 19402–19411.

Koelle, M.R., 2006. Heterotrimeric G protein signaling: getting inside the cell. Cell 126, 25–27.

Koelle, M.R., 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.

Koonin, E.V., Aravind, L., 2000. Dynein light chains of the Roadblock/LC7 group belong to an ancient protein superfamily implicated in NTPase regulation. Curr. Biol. 10, R774–R776.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

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

Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., Sigler, P.B., 1996. The 2.0 A crystal structure of a heterotrimeric G protein. Nature 379, 311–319.

Lassmann, T., Sonnhammer, E.L., 2005. Kalign—an accurate and fast multiple sequence alignment algorithm. BMC Bioinform. 6, 298.

Lee, C., Goldberg, J., 2010. Structure of coatomer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats. Cell 142, 123–132.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. Classification and evolution of P-loop GTPases and related ATPases. J. Mol. Biol. 317, 41–72.

Li, Y., Kelly, W.G., Logsdon Jr., J.M., Schurko, A.M., Harfe, B.D., Hill-Harfe, K.L., Kahn, R.A., 2004. Functional genomic analysis of the ADP-ribosylation factor family of GTPases: phylogeny among diverse eukaryotes and function in *C. elegans*. FASEB J. 18, 1834–1850.

Maeda, Y., Ide, T., Koike, M., Uchiyama, Y., Kinoshita, T., 2008. GPHR is a novel anion channel critical for acidification and functions of the Golgi apparatus. Nat. Cell Biol. 10, 1135–1145.

McCoy, K.L., Hepler, J.R., 2009. Chapter 3 regulators of g protein signaling proteins as central components of g protein-coupled receptor signaling complexes. Prog. Mol. Biol. Transl. Sci. 86, 49–74.

Mou, T.C., Masada, N., Cooper, D.M., Sprang, S.R., 2009. Structural basis for inhibition of mammalian adenylyl cyclase by calcium. Biochemistry 48, 3387–3397.

Mougel, C., Zhulin, I.B., 2001. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem. Sci. 26, 582–584.

Natarajan, K., Ashley, C.A., Hadwiger, J.A., 2000. Related Galpha subunits play opposing roles during Dictyostelium development. Differentiation 66, 136–146.

Neuwald, A.F., 2007. Galpha Gbetagamma dissociation may be due to retraction of a buried lysine and disruption of an aromatic cluster by a GTP-sensing Arg Trp pair. Protein Sci. 16, 2570–2577.

Neves, S.R., Ram, P.T., Iyengar, R., 2002. G protein pathways. Science 296, 1636–1639.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int. J. Neural Syst. 8, 581–599.

Noel, J.P., Hamm, H.E., Sigler, P.B., 1993. The 2.2 A crystal structure of transducin-alpha complexed with GTP gamma S. Nature 366, 654–663.

Oka, Y., Korsching, S.I., 2009. The fifth element in animal Galpha protein evolution. Commun. Integr. Biol. 2, 227–229.

Pandey, S., Monshausen, G.B., Ding, L., Assmann, S.M., 2008. Regulation of root-wave response by extra large and conventional G proteins in Arabidopsis thaliana. Plant J. 55, 311–322.

Pandey, S., Nelson, D.C., Assmann, S.M., 2009. Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. Cell 136, 136–148.

Pei, J., Sadreyev, R., Grishin, N.V., 2003. PCMA: fast and accurate multiple sequence alignment based on profile consistency. Bioinformatics 19, 427–428.

Pierce, K.L., Premont, R.T., Lefkowitz, R.J., 2002. Seven-transmembrane receptors. Nat. Rev. Mol. Cell Biol. 3, 639–650.

Ponting, C.P., Aravind, L., 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. Trends Biochem. Sci. 24, 130–132.

Popov, S., Yu, K., Kozasa, T., 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.

Porter, M.Y., Koelle, M.R., 2009. Chapter 2 insights into RGS protein function from studies in *Caenorhabditis elegans*. Prog. Mol. Biol. Transl. Sci. 86, 15–47.

Prabhu, Y., Eichinger, L., 2006. The Dictyostelium repertoire of seven transmembrane domain receptors. Eur. J. Cell Biol. 85, 937–946.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490.

Rodbell, M., Birnbaumer, L., Pohl, S.L., Krans, H.M., 1971. The glucagon-sensitive adenyl cyclase system in plasma membranes of rat liver. V. An obligatory role of guanylnucleotides in glucagon action. J. Biol. Chem. 246, 1877–1882.

Rompler, H., Staubert, C., Thor, D., Schulz, A., Hofreiter, M., Schoneberg, T., 2007. G protein-coupled time travel: evolutionary aspects of GPCR research. Mol. Interv. 7, 17–25.

Ross, E.M., Gilman, A.G., 1977. Resolution of some components of adenylate cyclase necessary for catalytic activity. J. Biol. Chem. 252, 6966–6969.

Roush, W., 1996. Regulating G protein signaling. Science 271, 1056–1058.

Roy, A.A., Lemberg, K.E., Chidiac, P., 2003. Recruitment of RGS2 and RGS4 to the plasma membrane by G proteins and receptors reflects functional interactions. Mol. Pharmacol. 64, 587–593.

Scheffzek, K., Ahmadian, M.R., Wittinghofer, A., 1998. GTPase-activating proteins: helping hands to complement an active site. Trends Biochem. Sci. 23, 257–262.

Sethakorn, N., Yau, D.M. and Dulin, N.O. Non-canonical functions of RGS proteins. *Cell Signal* **22**, pp. 1274-81. Q3

Siderovski, D.P., Hessel, A., Chung, S., Mak, T.W., Tyers, M., 1996. A new family of regulators of G-protein-coupled receptors? Curr. Biol. 6, 211–212.

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001

ARTICLE IN PRESS

Siderovski, D.P., Willard, F.S., 2005. The GAPs, GEFs, and GDIs of heterotrimeric G-protein alpha subunits. Int. J. Biol. Sci. 1, 51–66.
Sierra, D.A., Gilbert, D.J., Householder, D., Grishin, N.V., Yu, K., Ukidwe, P., Barker, S.A., He, W., Wensel, T.G., Otero, G., Brown, G., Copeland, N.G., Jenkins, N.A., Wilkie, T.M., 2002. Evolution of the regulators of G-protein signaling multigene family in mouse and human. Genomics 79, 177–185.
Simon, M.I., Strathmann, M.P., Gautam, N., 1991. Diversity of G proteins in signal transduction. Science 252, 802–808.
Simpson, A.G., Inagaki, Y., Roger, A.J., 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol. Biol. Evol. 23, 615–625.
Singh, P., Wang, B., Maeda, T., Palczewski, K., Tesmer, J.J., 2008. Structures of rhodopsin kinase in different ligand states reveal key elements involved in G protein-coupled receptor kinase activation. J. Biol. Chem. 283, 14053–14062.
Soding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248.
Soundararajan, M., Willard, F.S., Kimple, A.J., Turnbull, A.P., Ball, L.J., Schoch, G.A., Gileadi, C., Fedorov, O.Y., Dowler, E.F., Higman, V.A., Hutsell, S.Q., Sundstrom, M., Doyle, D.A., Siderovski, D.P., 2008. Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits. Proc. Natl Acad. Sci. USA 105, 6457–6462.
Spink, K.E., Polakis, P., Weis, W.I., 2000. Structural basis of the Axin-adenomatous polyposis coli interaction. EMBO J. 19, 2270–2279.
Srinivasa, S.P., Bernstein, L.S., Blumer, K.J., Linder, M.E., 1998. Plasma membrane localization is required for RGS4 function in Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA 95, 5584–5589.
Sun, B., Firtel, R.A., 2003. A regulator of G protein signaling-containing kinase is important for chemotaxis and multicellular development in dictyostelium. Mol. Biol. Cell 14, 1727–1743.
Sunahara, R.K., Taussig, R., 2002. Isoforms of mammalian adenylyl cyclase: multiplicities of signaling. Mol. Interv. 2, 168–184.
Tesmer, J.J., 2009. Chapter 4 structure and function of regulator of g protein signaling homology domains. Prog. Mol. Biol. Transl. Sci. 86, 75–113.
Tesmer, J.J., Berman, D.M., Gilman, A.G., Sprang, S.R., 1997. Structure of RGS4 bound to AlF4–activated G(i alpha1): stabilization of the transition state for GTP hydrolysis. Cell 89, 251–261.
Tesmer, J.J., Tesmer, V.M., Lodowski, D.T., Steinhagen, H., Huber, J., 2010. Structure of human G protein-coupled receptor kinase 2 in complex with the kinase inhibitor balanol. J. Med. Chem. 53, 1867–1870.
Tesmer, V.M., Kawano, T., Shankaranarayanan, A., Kozasa, T., Tesmer, J.J., 2005. Snapshot of activated G proteins at the membrane: the Galphaq-GRK2-Gbetagamma complex. Science 310, 1686–1690.
Tu, Y., Woodson, J., Ross, E.M., 2001. Binding of regulator of G protein signaling (RGS) proteins to phospholipid bilayers. Contribution of location and/or orientation to Gtpase-activating protein activity. J. Biol. Chem. 276, 20160–20166.
Versele, M., Lemaire, K., Thevelein, J.M., 2001. Sex and sugar in yeast: two distinct GPCR systems. EMBO Rep. 2, 574–579.
Wensel, T.G., 2008. Signal transducing membrane complexes of photoreceptor outer segments. Vis. Res. 48, 2052–2061.
Yeagle, P.L., Albert, A.D., 2007. G-protein coupled receptor structure. Biochim. Biophys. Acta 1768, 808–824.
Yona, S., Lin, H.H., Siu, W.O., Gordon, S., Stacey, M., 2008. Adhesion-GPCRs: emerging roles for novel receptors. Trends Biochem. Sci. 33, 491–500.
Yu, J.H., 2006. Heterotrimeric G protein signaling and RGSs in *Aspergillus nidulans*. J. Microbiol. 44, 145–154.
Zhu, H., Li, G.J., Ding, L., Cui, X., Berg, H., Assmann, S.M., Xia, Y., 2009. Arabidopsis extra large G-protein 2 (XLG2) interacts with the Gbeta subunit of heterotrimeric G protein and functions in disease resistance. Mol. Plant 2, 513–525.
Zhulin, I.B., Nikolskaya, A.N., Galperin, M.Y., 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J. Bacteriol. 185, 285–294.

UNCORRECTED PROOF

Please cite this article as: Anantharaman, V., et al., Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, Gene (2010), doi:10.1016/j.gene.2010.12.001