Molecular Evolutionary Analysis of Vertebrate Transducins: A Role for Amino Acid Variation in Photoreceptor Deactivation

Yi G. Lin, Cameron J. Weadick, Francesco Santini & Belinda S. W. Chang

Springer

Springer

ORIGINAL ARTICLE

Molecular Evolutionary Analysis of Vertebrate Transducins: A Role for Amino Acid Variation in Photoreceptor Deactivation
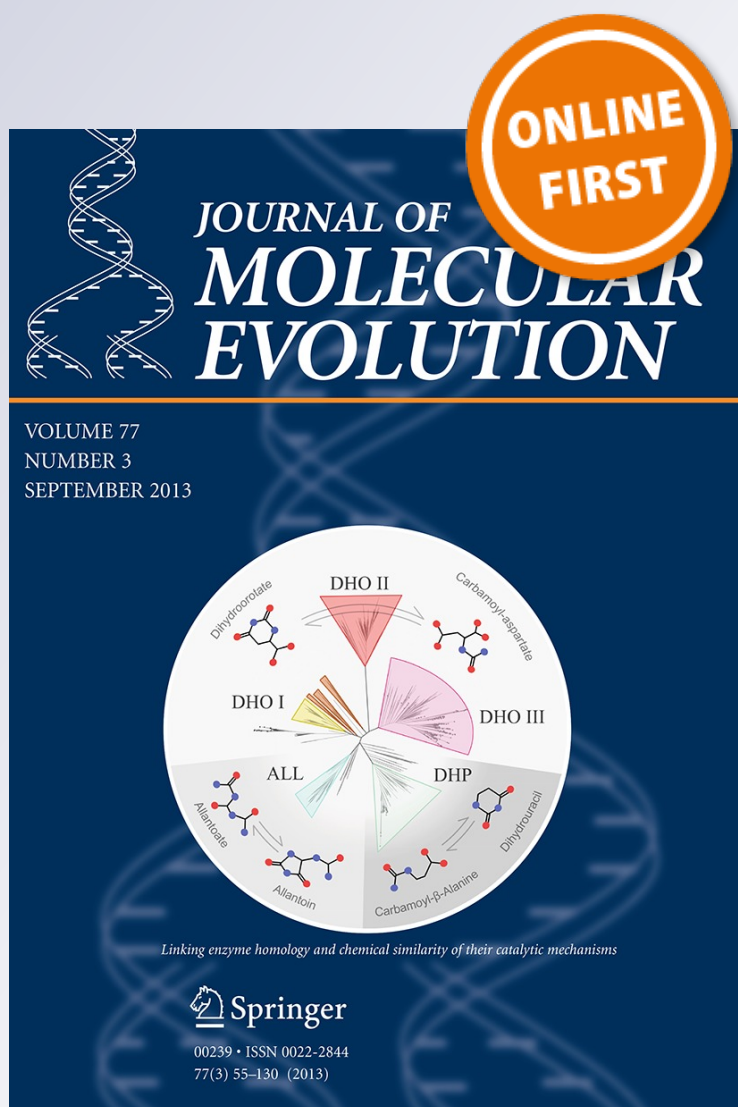
Yi G. Lin · Cameron J. Weadick · Francesco Santini · Belinda S. W. Chang

Abstract Transducin is a heterotrimeric G protein that plays a critical role in phototransduction in the rod and cone photoreceptor cells of the vertebrate retina. Rods, highly sensitive cells that recover from photoactivation slowly, underlie dim-light vision, whereas cones are less sensitive, recover more quickly, and underlie bright-light vision. Transducin deactivation is a critical step in photoreceptor recovery and may underlie the functional distinction between rods and cones. Rods and cones possess distinct transducin α subunits, yet they share a common deactivation mechanism, the GTPase activating protein (GAP) complex. Here, we used codon models to examine patterns of sequence evolution in rod (GNAT1) and cone (GNAT2) α subunits. Our results indicate that purifying selection is the dominant force shaping GNAT1 and GNAT2 evolution, but that GNAT2 has additionally been subject to positive selection operating at multiple phylogenetic scales; phylogeny-wide analysis identified several sites in the GNAT2 helical domain as having substantially elevated dN/dS estimates, and branch-site analysis identified several nearby sites as targets of strong positive selection during early vertebrate history. Examination of aligned GNAT and GAP complex crystal structures revealed steric clashes between several positively selected sites and the deactivating GAP complex. This suggests that GNAT2 sequence variation could play an important role in adaptive evolution of the vertebrate visual system via effects on photoreceptor deactivation kinetics and provides an alternative perspective to previous work that focused instead on the effect of GAP complex concentration. Our findings thus further the understanding of the molecular biology, physiology, and evolution of vertebrate visual systems.

Keyword G proteins · Vision · Rhodopsin · dN/dS · Positive selection · Maximum likelihood · Codon models

Y. Gloria Lin and Cameron J. Weadick have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-013-9589-5) contains supplementary material, which is available to authorized users.

Y. G. Lin · C. J. Weadick · F. Santini · B. S. W. Chang (✉)
Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada
e-mail: belinda.chang@utoronto.ca

Y. G. Lin · B. S. W. Chang
Department of Cell and Systems Biology, University of Toronto, Toronto, Canada

B. S. W. Chang
Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Canada

Introduction

The two primary photoreceptors of the vertebrate retina, the rods and cones, mediate vision under dim and bright conditions, respectively. Rods are more sensitive than cones, but they deactivate much more slowly; consequently, rods saturate under moderate light levels and trade off sensitivity with temporal resolution (reviewed in Ebrey and Koutalos 2001; Walls 1942). Considerable efforts have been directed toward understanding the molecular and biophysical processes that mediate rod and cone functions (reviewed in Fain et al. 2010; Yau and Hardie 2009). Visual phototransduction, the process by which light is converted into the biochemical signals that ultimately underlie the visual sense, begins with photon absorption by visual pigments in the retinal photoreceptors. In vertebrates, photoactivated visual

pigment binds and activates the α subunit of transducin (GNAT), inducing GDP-GTP exchange and the dissociation of GNAT from the transducin β and γ subunits. Activated GNAT then binds the inhibitory γ subunit of membrane-bound phosphodiesterase (PDE), permitting cGMP hydrolysis and, consequently, the closing of cGMP-gated cation channels in the plasma membrane, membrane hyperpolarization, and neurotransmitter release. The deactivation of photoactivated visual pigments and the increased synthesis of cGMP both contribute to the overall process of photoreceptor deactivation, but deactivation of the GNAT appears to be the rate-limiting step in rod response recovery (reviewed in Burns and Pugh 2010). GNAT deactivates via intrinsic GTPase activity, but this process is quite slow, and a GTPase activating protein (GAP) complex is required to elevate GNAT GTPase activity to physiologically relevant levels (He et al. 1998; Hu and Wensel 2002; Makino et al. 1999).

Rods and cones possess several unique protein variants adapted to their specific roles in rod or cone phototransduction, such as the rod and cone visual pigment proteins (opsins) (Nordstrom et al. 2004). Some phototransduction components, however, are shared between rod and cone photoreceptors. Notably, there are two copies of the transducin α subunit, one for rods (GNAT1) and one for cones (GNAT2) (Lerea et al. 1986; Lochrie et al. 1985b), but the rod and cone photoreceptor cells appear to share a single GAP complex (e.g., Cowan et al. 1998; Lyubarsky et al. 2001; Zhang et al. 2003). Inactivation of divergent α transducin subunits thus depends on the same complex in the functionally distinct rods and cones. As cone photoreceptor cells have been shown to possess higher concentrations of the deactivating GAP complex with respect to rods (Cowan et al. 1998; Lyubarsky et al. 2001; Zhang et al. 2003), the faster deactivation of cone photoreceptors may simply reflect more frequent interactions between GNAT2 and the GAP complex (Burns and Pugh 2009). However, the rate of rod deactivation is substantially faster in transgenic mice expressing cone GNAT2 in place of GNAT1 in the rod photoreceptor cells (Chen et al. 2010), and this suggests that differences in protein–protein interactions may also play an important role in explaining variation in photoreceptor recovery rates.

The structural basis of the interaction between GNAT and the GAP complex has been the subject of many years of study. It is known that GNAT consists of two domains, a GTPase domain and an α helical domain, between which lies the nucleotide binding cleft; following GTP hydrolysis and GNAT activation, several regions ('switches') of the GTPase domain undergo conformational changes that facilitate interactions with both PDE and the GAP complex (reviewed in Oldham and Hamm 2008; Wittinghofer 1994). Within the GAP complex, the regulator of G protein signaling (RGS) domain of the RGS9 protein is known to play a central role in regulating GNAT GTPase activity (He et al. 1998; Hu and Wensel 2002; Makino et al. 1999). The 3D structures of several components of the GNAT–GAP complex interaction have been solved, providing important insights into the structural and functional roles played by different regions of GNAT and the proteins of the GAP complex (Cheever et al. 2008; Lambright et al. 1994; Noel et al. 1993; Slep et al. 2001; Sondek et al. 1994). These studies can also provide a framework for evolutionary analysis of GNAT protein structure and function.

Though rod and cone phototransduction proteins serve well-defined functions in vision, the specific selection pressures that these two sets of proteins experience will likely vary among species depending on ecological requirements or physiological constraints. Perhaps most importantly, there can be considerable variation in ambient light intensity across environments (e.g., between aquatic and terrestrial or nocturnal and diurnal niches) (Lythgoe 1979). Because the interaction between GNAT and the GAP complex is critical to photoreceptor recovery rate, which affects key visual parameters such as absolute sensitivity and temporal resolution, we carried out a comparative, molecular evolutionary study of vertebrate rod and cone GNAT sequences. Working with data sets composed of newly derived and previously determined rod and cone GNAT sequences, we applied codon substitution models that, by estimating the nonsynonymous to synonymous substitution rate ratio (dN/dS or ω), provide insight into the nature of selection operating on protein coding DNA (reviewed in Anisimova and Liberles 2012). We focused our sequencing efforts on ray-finned fishes, as it is a speciose and ecologically diverse group that has been understudied at the level of whole genomes (from which we obtained much of our GNAT data). Our results suggest that positive selection has shaped GNAT2 evolution at multiple evolutionary scales, likely in ways that affect the interaction between GNAT2 and the deactivating GAP complex.

Methods

GNAT1 and GNAT2 Data Sets

Data sets of 39 GNAT1 and 36 GNAT2 sequences were obtained through a combination of (1) online database searches and (2) targeted PCR and DNA sequencing. Species names and accession numbers for the analyzed sequences are provided in Supplementary Table 1. Previously determined GNAT1 and GNAT2 sequences were obtained from NCBI Genbank and Ensembl via sequence similarity and keyword searches; in a few cases, predicted/

unannotated sequences were extracted from ongoing genome project databases. Degenerate PCR primers were designed to match conserved gene regions following preliminary alignments of these downloaded sequences; primer names and sequences are provided in Supplementary Table 2. These primers were used to amplify GNAT1 and GNAT2 sequences from eye cDNA libraries for 16 species of ray-finned fish; all were obtained through the pet trade save the Trinidadian pike cichlid (*Crenicichla frenata*), which was wild-caught. Fish were killed following sedation via ice baths or immersion in tricaine methanesulfonate solution, after which eyes were either excised and stored at −80 °C or preserved in RNALater stabilization solution (Roche). cDNA libraries, prepared using the SMART cDNA Library Construction Kit (BD Biosciences), were used as template for degenerate PCR using standard chemistry and thermal-cycling settings, with the FastStart Taq polymerase (Roche). Purified PCR products were ligated into TA-overhang plasmid vectors (Invitrogen), cultured in *Escherichia coli*, and multiple clones sequenced using vector primers.

Alignments

GNAT1 and GNAT2 data sets were initially compiled and aligned using ClustalW (Thompson et al. 1994) as implemented in MEGA 4 (Tamura et al. 2007). Nucleotide sequences were translated to amino acids prior to alignment. As alignment errors can lead to false signature of positive selection (Fletcher and Yang 2010; Jordan and Goldman 2012), we generated additional alignments using PRANK (Loytynoja and Goldman 2005) and PAGAN (Loytynoja et al. 2012), alignment methods that outperform ClustalW over a wide range of divergence levels and indel rates (Fletcher and Yang 2010; Jordan and Goldman 2012). As before, alignments were conducted using amino acid translated sequences. We then estimated confidence in these PRANK and PAGAN alignments in GUIDANCE, which uses a bootstrapping approach to perturb the alignment guide tree (Penn et al. 2010a, b). Manual inspection of the alignments revealed that results were identical regardless of algorithm for both data sets. Moreover, GUIDANCE confidence scores were effectively maximized (PRANK: 1.000 for both GNAT1 and GNAT2; PAGAN: 0.999 for GNAT1 and 0.998 for GNAT2). Translated alignments are displayed in Supplementary Figures 1 and 2. The GNAT1 and GNAT2 alignments were each 187 codons long, spanning amino acid (AA) positions 67–253 (AA numbering corresponds to bovine GNAT1; Genbank ID NP_851365) with only a few of the sequences truncated at the N- and/or C-terminal ends of the alignment. There were no internal gaps in the GNAT1 alignment and only one internal gap in the GNAT2 alignment (a one codon deletion in the *Tetraodon nigroviridis*

GNAT2 sequence at AA position 73). The GNAT1 and GNAT2 data sets were nearly but not entirely taxonomically equivalent (a consequence of variation in sequence availability for each species); particular analyses were therefore repeated using reduced data sets consisting of comparable taxonomic sets.

Selection Analyses

In order to explore patterns of selective constraint across alignments and/or phylogenies of GNAT1 and GNAT2 genes, a variety of codon substitution models were fit to the aligned data sets using the CODEML program of the PAML software package (Yang 2007). Codon substitution models provide estimates of the nonsynonymous (dN) to synonymous (dS) substitution rate ratio (ω or dN/dS), which describes the form and strength of selection operating on protein coding DNA; $\omega > 1$ is indicative of positive selection, $\omega = 1$ is consistent with neutrality, and $\omega < 1$ reflects purifying selection (Goldman and Yang 1994; Muse and Gaut 1994). Different codon substitution models make different assumptions about whether and how the ω ratio varies across the alignment and/or the phylogeny (described below). The relative goodness-of-fit of particular models was established by comparing complex models against simpler, nested models via likelihood ratio tests (LRTs), with twice the difference in log likelihood scores compared against a χ^2 null distribution (Huelsenbeck and Rannala 1997); unless otherwise noted, the degrees of freedom for each LRT equaled the number of additional parameters estimated in the more complicated, alternative model. Recent estimates of species-level relationships were consulted to describe evolutionary relationships among the GNAT sequences for the rod and cone transducin data sets (Maddison et al. 2007).

One common concern when analyzing highly divergent sequences is the reliability of the parameter estimates (dN/dS, distances, etc.) (Gojobori 1983; Maynard Smith 1994), as synonymous substitutions may saturate over long timescales and as extremely divergent sequences are often difficult to align accurately. Fortunately, our alignments appear to be sound (described above), and simulation results have shown that the use of well-chosen codon models and maximum likelihood (ML) estimation methods generally result in accurate parameter estimates even when sequences are highly divergent as long as the alignment is correct (Anisimova et al. 2001; Fletcher and Yang 2010; Gharib and Robinson-Rechavi 2013; Jordan and Goldman 2012). Moreover, by employing a phylogenetic rather than pairwise approach, we avoid directly estimating evolutionary distances between extremely divergent sequences; instead, branches leading to internodes and other sequences repeatedly bisect these long distances. To confirm that synonymous sites are not saturated

in our data sets, we conducted a preliminary investigation using the simple M0 codon substitution model, which assumes that ω is invariant across both sites and branches (Goldman and Yang 1994). These results suggest that our data sets are not strongly affected by synonymous saturation. For GNAT1 ($\omega = 0.023$, $\kappa = 2.178$, $\ln L = -6,509.116$, number of parameters $= 73$), branch-specific estimates of dS (the number of synonymous substitutions per synonymous site) ranged from 0.00 to 1.57, with a median of 0.18. For GNAT2 ($\omega = 0.062$, $\kappa = 1.815$, $\ln L = -7,019.283$, number of parameters $= 67$), branch-specific estimates of dS ranged from 0.00 to 0.90, with a median of 0.17. Critically, dS values were well below 1 along the branches connecting the fish and tetrapod clades in the two data sets.

Codon Models

Random-sites models allow for variation in ω among sites in the alignment but not among branches of the phylogeny. Two different random-sites LRTs were employed to test for site-specific positive selection: M8 versus M7 and M8 versus M8a (Swanson et al. 2003; Yang et al. 2000). The M8, M8a, and M7 models accommodate among-site ω variation across the range $0 < \omega < 1$ by using a continuous probability distribution (a β distribution); M8a and M8 expand on the simpler M7 model by including either a neutral site class ($\omega_2 = 1$) or a positively selected site class ($\omega_2 > 1$), respectively. For each model, ML was used to estimate the ω ratio parameters (including p and q, the β distribution shape parameters), the sizes of the site classes (data set proportions; p), the transition-to-transversion nucleotide substitution rate ratio (κ), and branch lengths. Stationary codon frequencies were estimated using the F3×4 approximation method. We opted to use the F3×4 method, rather than the Fcodon method, as the latter approach would involve the estimation of 61 sense codon frequencies from an alignment only 187 codon long. However, we note that codon frequencies estimated under the two methods were positively correlated (GNAT1: Spearman's $\rho = 0.66$, $P < 0.001$; GNAT2: Spearman's $\rho = 0.58$, $P < 0.001$). In order to detect and avoid local optima, models were fit to the data sets multiple times from different initial parameter values. Alignment gaps were treated as ambiguous characters and retained in analyses.

Branch-site and clade models allow for variation in ω both among sites and among branches, with variation accommodated through the a priori partitioning of the phylogeny into 'foreground' and 'background' lineages (Yang and Nielsen 2002). Numerous ecological and physiological factors could conceivably result in lineage-specific patterns of GNAT sequence evolution (e.g., diurnal activity pattern, physiological temperature); here, partitioning schemes were designed to contrast evolutionary patterns within the primarily terrestrial tetrapod clade and the aquatic ray-finned fish clade.

Branch-site and clade models were carried out as described above for random-site models. For clade models, we employed the model C approach (CmC) of Bielawski and Yang (2004), the multi-clade extension of Yoshida et al. (2011), and the null model (M2a_rel) of Weadick and Chang (2012). CmC assumes three site classes, which account for pervasive purifying selection ($0 < \omega_0 < 1$), pervasive neutrality ($\omega_1 = 1$), and divergent selection ($\omega_2 > 0$ and $\omega_3 > 0$ for two prespecified partitions of a phylogeny). The M2a_rel null model is similarly constructed, but removes divergent selection through the constraint $\omega_2 = \omega_3$. The multi-clade CmC extension allows for the designation of multiple foreground partitions ($\omega_2 > 0$, $\omega_3 > 0$, $\omega_4 > 0$, etc.). Five models were fit to each data set (Fig. 1): (1) a multi-clade CmC with separate partitions for the fish clade, the tetrapod clade, and the ancestral branch connecting the two clades; (2, 3) simpler clade models that united the ancestral branch with either the fish (CmC-v1) or tetrapod (CmC-v2) clade, respectively; (4) a simple clade model that united the fish and tetrapod clades in a common partition to the exclusion of the ancestral branch (CmC-v3); and (5) an M2a_rel null model.

The branch-site model A (test 2) approach of Zhang et al. (2005) was used to test for positive selection along the branch connecting the tetrapod and ray-finned fish GNAT clades. The alternative model for this test assumes four site classes: pervasive purifying selection ($0 < \omega_0 < 1$), pervasive neutrality ($\omega_1 = 1$), purifying-to-positive selection ($\omega_2 > 1$ along the foreground branch, $0 < \omega_0 < 1$ otherwise), and neutral-to-positive selection ($\omega_2 > 1$ along the foreground branch, $\omega_1 = 1$ otherwise). The null model is similarly constructed but does not permit positive selection ($\omega_2 = 1$). The branch-site model of Zhang et al. (2005) assumes that selection within the background partition of the phylogeny is constant across branches and not affected by positive selection; as such, the recently proposed Branch-site REL method of Kosakovsky Pond et al. (2011), which relaxes both of these assumptions, was employed as well (via the Datamonkey web server) (Delport et al. 2010). As with Zhang et al.'s method, the LRT for the branch-site REL method compares the fit of models that either do ($\omega_+ > 1$) or do not ($\omega_+ = 1$) allow for positive selection on the branch of interest. Finally, we also applied Zhang et al.'s (2005) branch-site test to entire clades (fishes or tetrapods) in a manner analogous to the clade models. As this model only allows for a single-foreground partition, and as the analyses are conducted on unrooted trees, analyses were conducted twice, with the ancestral branch assigned to either the foreground or background partition.

Structural Analyses

Following random-site and branch-site analyses, codons were identified as 'positively selected' through Bayes

Fig. 1 Phylogenetic partitioning scheme employed in clade model analysis of the GNAT1 and GNAT2 data sets. Three partitions were assumed: the tetrapod clade, the ray-finned fish clade, and the ancestral branch connecting the two clades. The models differ in whether and how ω (= dN/dS) diverges across the three partitions. All models possess similar purifying and neutral site classes, which apply to all phylogenetic partitions

		Clade model				
Site class		M2a_rel	CmC v1	CmC v2	CmC v3	Multiclade CmC
Purifying		$0 < \omega_0 < 1$	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$
Neutral		$\omega_1 = 1$	$\omega_1 = 1$	$\omega_1 = 1$	$\omega_1 = 1$	$\omega_1 = 1$
Divergent						
	Tetrapods	$\omega_2 > 0$	$\omega_2 > 0$	$\omega_2 > 0$	$\omega_2 > 0$	$\omega_2 > 0$
	Ancestral branch	$= \omega_2$	$\omega_3 > 0$	$= \omega_2$	$\omega_3 > 0$	$\omega_3 > 0$
	Fishes	$= \omega_2$	$= \omega_3$	$\omega_3 > 0$	$= \omega_2$	$\omega_4 > 0$

empirical Bayes (BEB) assignment of codons to site classes with $\omega_2 > 1$ (Yang et al. 2005). Ancestral amino acid character states were inferred by ML using CODEML (Yang et al. 1995) and assuming the JTT + G4 amino acid substitution model (Jones et al. 1992; Yang 1994), which was selected for use following preliminary investigation using ProtTest (Abascal et al. 2005). 3D structures of ancestral proteins were then inferred via homology modeling using Modeller (Sali and Blundell 1993); for each model, the Modeller objective function was minimized 25 separate times, and the run with the lowest DOPE score (Shen and Sali 2006) was selected for assessment and visualization. Model quality was first assessed using Pro-Check (Laskowski et al. 1993), which showed that overall G-factors were positive, indicating that main-chain bond lengths and main- and side-chain bond angles were generally in high-probability stereochemical conformations (results not shown). Model quality was then assessed using ProSA-web (Wiederstein and Sippl 2007), which showed that the model and template structures had comparable z-scores (a measure of how a structure's total energy compares to that expected by random chance), standardized for the number of residues (results not shown). The template structure for GNAT homology modeling was a chimeric G protein α subunit solved in complex with the RGS domain of bovine RGS9 (protein data bank ID 1FQJ) (Slep et al. 2001); this chimeric protein was based primarily on bovine GNAT1 (92 % identical), but with a small portion of the GTPase domain replaced with the equivalent region a homologous α subunit from the inhibitory G protein subfamily (GNAI). Using MacPyMol (Delano Scientific), the GNAT model was then docked onto a GAP complex by aligning this RGS domain with the equivalent portion of a solved crystal structure of RGS9 and GNB5 (protein data bank ID 2PBI), following Cheever et al. (2008). It should be noted that the interaction between these proteins could also be examined using more sophisticated approaches that explicitly account for factors such as residue charge/

polarity and backbone flexibility. Such analyses may provide more subtle insights into the protein–protein interactions within the GAP complex, but they are beyond the scope of the present paper.

Results

The results of random-sites analyses of the GNAT1 and GNAT2 data sets are provided in Table 1. Overall, estimated dN/dS was lower and less variable across alignment sites for GNAT1 than for GNAT2. For the GNAT1 data set, the M8 model failed to improve on either of its null models (M7 or M8a; $P = 1.000$ in both cases). The ML parameter estimates obtained under these models indicate that dN/dS was skewed strongly toward zero (β distribution parameter estimates: $p = 0.118$, $q = 2.446$, $p_0 = 1.000$), and no sites were assigned to the 'positively selected' site class by BEB analysis. For the GNAT2 data set, however, the M8 model did improve on the M7 null model ($P < 0.001$), and the LRT for the M8 versus M8a comparison approached significance ($P = 0.117$ assuming a χ_1^2 null distribution, or $P = 0.059$ assuming a 50:50 mixture distribution of 0 and a χ_1^2 distribution that accounts for a boundary condition in the LRT) (Goldman and Whelan 2000). Parameter estimates obtained under the M8 model suggest that dN/dS is less strongly skewed toward dN/dS = 0 than was observed for the GNAT1 data set ($p = 0.201$, $q = 3.502$, $p_0 = 0.979$, $\omega_2 = 1.257$). Moreover, parameter estimates indicated that a small fraction of the data set evolved with dN/dS > 1, with four sites identified by BEB analysis as members of the 'positively selected' site class (sites 94, 95, 96, and 117, given bovine GNAT1 amino acid numbering). These sites are all located in the protein's helical domain (sites 94–96 in helix B and site 117 in helix C). Site-by-site posterior mean estimates of dN/dS for GNAT1 and GNAT2 are provided in Fig. 2;

Table 1 Random-site results for the GNAT1 and GNAT2 data sets

Data set	Model (n.p.)[a]	$\ln L$[b]	Parameter estimates	LRT P values	Positively selected sites (PP)[c]
GNAT1	M7 (74)	−6,322.26	$\kappa = 2.213$, $p = 0.118$, $q = 2.446$	–	–
	M8a (75)	−6,322.26	$\kappa = 2.213$, $p_0 = 1.000$	–	–
			$p = 0.118$, $q = 2.447$		
	M8 (76)	−6,322.26	$\kappa = 2.213$, $p_0 = 1.000$, $\omega_2 = 1.000$	Versus M7: $P = 1.000$	None
			$p = 0.118$, $q = 2.447$	Versus M8a: $P = 1.000$	
GNAT2	M7 (68)	−6,706.93	$\kappa = 1.783$, $p = 0.142$, $q = 1.070$	–	–
	M8a (69)	−6,692.42	$\kappa = 1.782$, $p_0 = 0.978$	–	–
			$p = 0.201$, $q = 3.598$		
	M8 (70)	−6,691.19	$\kappa = 1.780$, $p_0 = 0.979$, $\omega_2 = 1.257$	Versus M7: $P < 0.001$	94 (0.99), 95 (0.58)
			$p = 0.201$, $q = 3.502$	Versus M8a: $P = 0.117$	96 (0.90), 117 (0.95)

[a] Refer to main text for explanation of models and tests. 'n.p.' = number of estimated parameters

[b] '$\ln L$' = log likelihood score

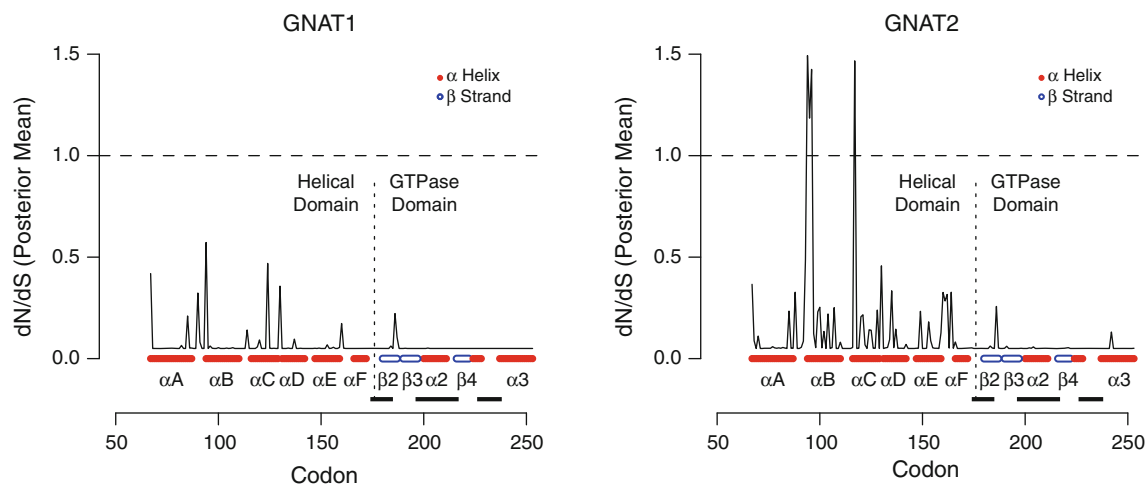[c] Site numbering follows that of bovine GNAT1. 'PP' = posterior probability

Fig. 2 Site-specific dN/dS estimates for the GNAT1 (*left*) and GNAT2 (*right*) data sets, obtained using the M8 random-site model. GNAT secondary structure for the analyzed fragment is noted above the *X*-axes (*solid bars* α helices, *outlined bars* β strands). The three 'switch' regions are noted with *black rectangles*. Site numbering follows bovine GNAT1

the higher average site-specific dN/dS for GNAT2 compared to GNAT1 and the 'positively selected' sites are both readily apparent. Furthermore, site-specific dN/dS estimates were, on average, higher in the helical domain than in the GTPase domain for both data sets. Qualitatively similar results were obtained from random-sites analyses of reduced data sets comprised only of species for which we possessed both GNAT1 and GNAT2 sequences (results not shown).

Ancestral GNAT2 amino acid sequences were inferred in order to reconstruct the evolutionary history of the sites identified as positively selected by random-site BEB analysis. Overall confidence in the reconstructions was high, with the posterior probability of the most probable residue averaging 0.994 across all sites and nodes; the across-node average for the four positively selected sites ranged from 0.937 to 0.986. Counting changes between adjacent node reconstructions within the GNAT2 phylogeny revealed an average of 1.7 changes per site (range 0–15; mode = 0). Over half of the sites in the helical domain changed at least once (68 of 109 sites), but only a small minority of sites within the GTPase domain likewise varied (11 of 78 sites). The four positively selected sites were highly variable, changing, at minimum, 11–15 times across the phylogeny, with 6–7 amino acid residues implicated throughout the history of each site (Fig. 3); ten or fewer changes were inferred for all other sites, save one (site 162, 12 changes, 5 residues). Note that these estimates of the number of substitutions at each site are likely underestimates owing to repeated substitutions at a single site. Considering the four positively selected sites, substitutions were inferred along 32 of the GNAT2 phylogeny's branches (12 within the

Fig. 3 GNAT2 cladogram showing the inferred substitution history for each of the four 'positively selected' sites identified by random-site BEB analysis. *Boxes* indicate AA substitutions; the four AA positions are distinguished by box outline, and the ancestral (*top*) and derived (*bottom*) residue states are indicated using one-letter AA codes (see *inset*). Note that the ancestral versus derived relationship is ambiguous for changes on the unrooted branch that connects the tetrapod (*top*) and fish (*bottom*) clades. Site numbering follows that of bovine GNAT1
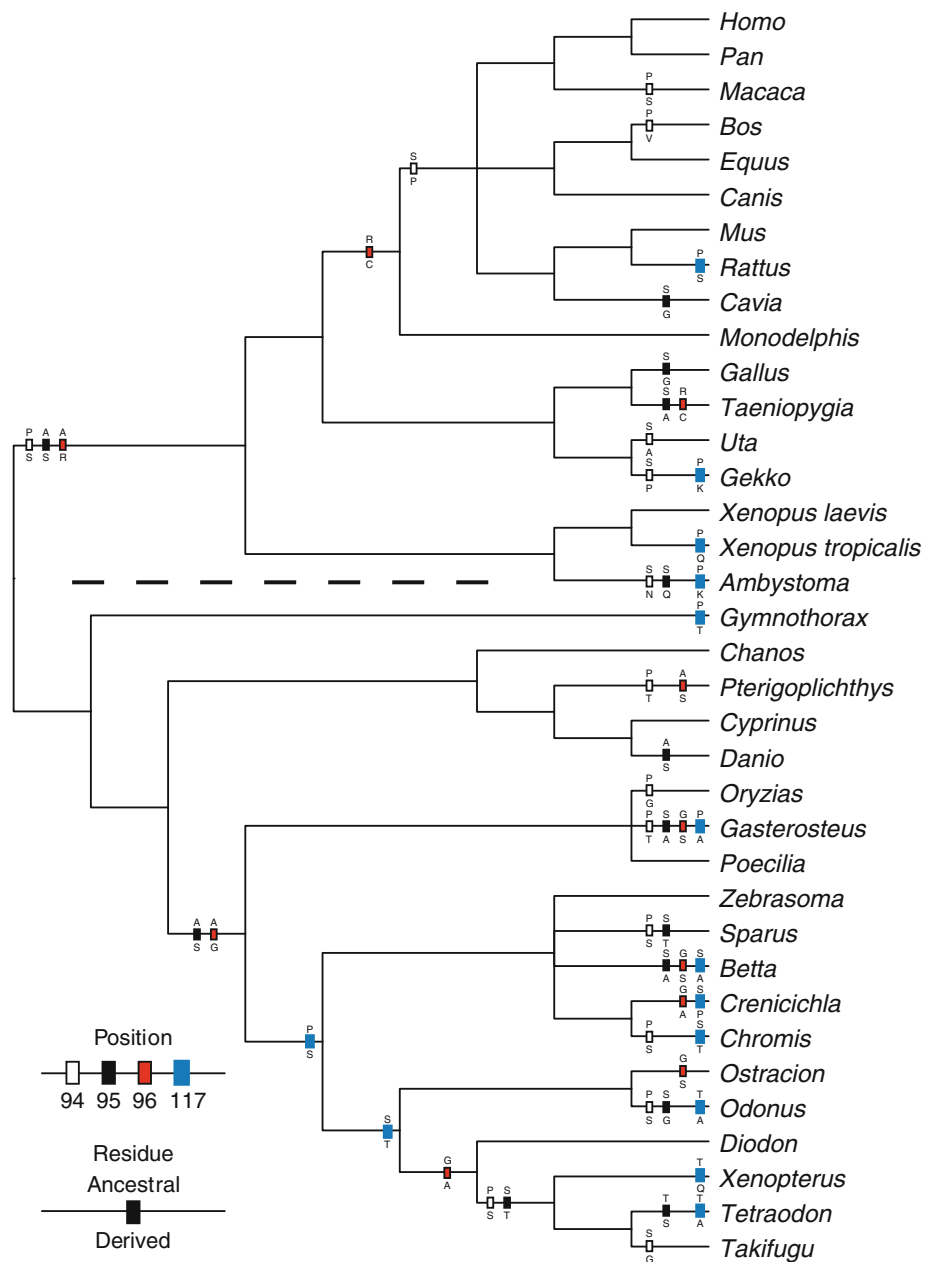
tetrapod clade, 19 within the fish clade, plus the ancestral branch uniting the two clades), and co-occurring changes at more than one of these sites were observed along 14 branches (Fig. 3).

The results of clade model analyses of the GNAT1 and GNAT2 data sets are provided in Tables 2 and 3, respectively. A few general patterns emerged from these analyses. For both data sets and regardless of partitioning strategy, parameter estimates indicated that dN/dS was very low ($\omega_0 \approx 0$) for the majority of the data ($p_0 > 0.7$). Furthermore, in both cases, the multi-clade model significantly improved on the M2a_rel null model (LRT $P < 0.01$ in each case), with parameter estimates indicating that dN/

dS was highest along the ancestral branch and lowest within the tetrapod partition. However, additional LRTs revealed differences between the two data sets in the effect of particular partitioning schemes. For GNAT1, treating the fish and tetrapod clades as separate partitions significantly improved model fit, and this was true regardless of whether an additional partition was included for the ancestral connecting branch (all $P < 0.01$; LRTs: multiclade CmC vs. CmC-v3, CmC-v1 vs. M2a_rel, and CmC-v2 vs. M2a_rel). Including an additional partition for the ancestral connecting branch did not significantly improve model fit, regardless of whether the fish and tetrapod clades were united or considered as independent partitions (all

Table 2 Clade model C (CmC) results for the GNAT1 data set

Model (n.p.)[a]	$\ln L$	Parameter estimates	LRT P values
Multi-clade (78)	$-6{,}327.68$	$\kappa = 2.303$, $p_0 = 0.805$, $p_1 = 0.013$	Versus M2a_rel: $P = 0.007$,
Partitions		$\omega_0 = 0.003$	Versus CmC-v1: $P = 0.225$
RFF, T, AB		$\omega_{RFF} = 0.130$, $\omega_T = 0.069$, $\omega_{AB} = 0.476$	Versus CmC-v2: $P = 0.085$, Versus CmC-v3: $P = 0.004$
CmC-v1 (77)	$-6{,}328.42$	$\kappa = 2.309$, $p_0 = 0.802$, $p_1 = 0.015$	Versus M2a_rel: $P = 0.003$
Partitions		$\omega_0 = 0.003$	
RFF + AB, T		$\omega_{RFF+AB} = 0.129$, $\omega_T = 0.069$	
CmC-v2 (77)	$-6{,}329.17$	$\kappa = 2.305$, $p_0 = 0.801$, $p_1 = 0.016$	Versus M2a_rel: $P = 0.008$
Partitions		$\omega_0 = 0.003$	
T + AB, RFF		$\omega_{T+AB} = 0.071$, $\omega_{RFF} = 0.125$	
CmC-v3 (77)	$-6{,}331.90$	$\kappa = 2.257$, $p_0 = 0.802$, $p_1 = 0.017$	Versus M2a_rel: $P = 0.202$
Partitions		$\omega_0 = 0.003$	
T + RFF, AB		$\omega_{T+RFF} = 0.100$, $\omega_{AB} = 0.392$	
M2a_rel (76)	$-6{,}332.71$	$\kappa = 2.265$, $p_0 = 0.801$, $p_1 = 0.017$	–
Partitions		$\omega_0 = 0.003$	
N/A		$\omega_2 = 0.101$	

T tetrapod clade, RFF ray-finned fish clade, AB ancestral branch. Refer to Table 1 for additional definitions

[a] Refer to main text and Fig. 1 for explanation of models and tests

Table 3 CmC results for the GNAT2 data set

Model (n.p.)[a]	$\ln L$	Parameter estimates	LRT P values
Multi-clade (72)	$-6{,}689.11$	$\kappa = 1.787$, $p_0 = 0.695$, $p_1 = 0.022$	Versus M2a_rel: $P < 0.001$,
Partitions		$\omega_0 = 0.005$	Versus CmC-v1: $P = 0.001$
RFF, T, AB		$\omega_{RFF} = 0.192$, $\omega_T = 0.1300$, $\omega_{AB} = 0.976$	Versus CmC-v2: $P < 0.001$, Versus CmC-v3: $P = 0.012$
CmC-v1 (71)	$-6{,}694.81$	$\kappa = 1.800$, $p_0 = 0.704$, $p_1 = 0.024$	Versus M2a_rel: $P = 0.002$
Partitions		$\omega_0 = 0.006$	
RFF + AB, T		$\omega_{RFF+AB} = 0.213$, $\omega_T = 0.134$	
CmC-v2 (71)	$-6{,}697.69$	$\kappa = 1.800$, $p_0 = 0.705$, $p_1 = 0.025$	Versus M2a_rel: $P = 0.059$
Partitions		$\omega_0 = 0.006$	
T + AB, RFF		$\omega_{T+AB} = 0.149$, $\omega_{RFF} = 0.199$	
CmC-v3 (71)	$-6{,}692.27$	$\kappa = 1.777$, $p_0 = 0.697$, $p_1 = 0.022$	Versus M2a_rel: $P < 0.001$
Partitions		$\omega_0 = 0.005$	
T + RFF, AB		$\omega_{T+RFF} = 0.161$, $\omega_{AB} = 0.972$	
M2a_rel (70)	$-6{,}699.47$	$\kappa = 1.791$, $p_0 = 0.703$, $p_1 = 0.025$	–
Partitions		$\omega_0 = 0.006$	
N/A		$\omega_2 = 0.173$	

T tetrapod clade, RFF ray-finned fish clade, AB ancestral branch. Refer to Table 1 for additional definitions

[a] Refer to main text and Fig. 1 for explanation of models and tests

$P > 0.08$; LRTs: multi-clade CmC vs. CmC-v1, multi-clade CmC vs. CmC-v2, and CmC-v3 vs. M2a_rel). For GNAT2, as for GNAT1, treating the fish and tetrapod clades as separate partitions led to significant (or nearly significant) improvements in model fit (all $P < 0.06$; LRTs: multi-clade CmC vs. CmC-v3, CmC-v1 vs. M2a_rel, and CmC-v2 vs. M2a_rel). However, contrary to what was observed for GNAT1, including an additional partition for the ancestral connecting branch did significantly improve model fit (all $P < 0.01$; LRTs: multi-clade CmC vs. CmC-v1, multi-clade CmC vs. CmC-v2, and CmC-v3 vs. M2a_rel). The clade model analyses thus indicated (1) that dN/dS was higher in the fish clade than in the tetrapod clade for both the GNAT1 data set and the

Table 4 Branch-site results for the GNAT1 and GNAT2 data sets

Data set	Model (n.p.)[a]	lnL	Parameter estimates	P value	Positively selected sites (PP)
GNAT1	Alt (76)	−6,432.55	$\kappa = 2.320$, $p_0 = 0.922, p_1 = 0.033, \omega_0 = 0.017, \omega_2 = 1.000$	P = 1.000	None
	Null (75)	−6,432.55	$\kappa = 2.320$, $p_0 = 0.922, p_1 = 0.033, \omega_0 = 0.017$	–	–
GNAT2	Alt (70)	−6,839.41	$\kappa = 1.923$, $p_0 = 0.829, p_1 = 0.065, \omega_0 = 0.038, \omega_2 = 11.697$	P = 0.0022	77 (0.98), 85 (1.00), 91 (0.76), 92 (0.61), 93 (0.51), 96 (0.55), 97 (0.66), 101 (0.83), 102 (0.79), 104 (0.90), 120 (0.70), 130 (0.69), 135 (0.89), 166 (0.73), 183 (1.00), 187 (0.69), 242 (1.00), 253 (0.71)
	Null (69)	−6,844.08	$\kappa = 1.902$, $p_0 = 0.794, p_1 = 0.064, \omega_0 = 0.038$	–	–

[a] Refer to main text for explanation of models and tests. Refer to Table 1 for additional definitions

GNAT2 data set and (2) that dN/dS was significantly elevated for the GNAT2 ancestral branch but not for the GNAT1 ancestral branch.

Branch-site analyses confirmed that dN/dS was elevated along the ancestral branch in the GNAT2 data set, but yielded mixed evidence for the GNAT1 data set (Table 4). For GNAT2, the branch-site alternative model of Zhang et al. (2005) significantly improved on its null model (P = 0.0022), with the ML parameter estimates indicating that approximately 10 % of the data set evolved with dN/dS ≫ 1 along the ancestral branch ($p_2 = 0.106$, $\omega_2 = 11.697$). Using the branch-site REL method of Kosakovsky Pond et al. (2011) yielded qualitatively similar results ($\omega_+ = 17.76$, $p_+ = 8.0$ %, $P < 0.001$), and results from both methods were effectively unchanged if we instead analyzed the taxonomically reduced GNAT2 data set described above (data not shown). By contrast, results from similar analyses of the GNAT1 data set were inconsistent. For GNAT1, the branch-site alternative model of Zhang et al. (2005) failed to improve on its null model when the full data set was used (P = 1.000), but approached significance when the taxonomically reduced data set was considered (P = 0.086 assuming a χ_1^2 null distribution, or P = 0.043 assuming a 50:50 mixture distribution of 0 and a χ_1^2 distribution that accounts for a boundary condition in the LRT) (Goldman and Whelan 2000). Similar results were obtained using Kosakovsky Pond et al.'s (2011) branch-site REL approach; the LRT was significant when the reduced data set was analyzed (P = 0.001) but not when the full data set was analyzed (P = 1.000). As some of the GNAT1 sequences removed from the full data set derived from basal fish lineages, their removal slightly altered the phylogenetic interpretation of the ancestral branch connecting the tetrapod and fish clades. Specifically, in the reduced data set, the ancestral branch was

longer and connected tetrapods to a neopterygiian fish clade rather than a larger actinopterygiian fish clade. This may indicate that patterns of GNAT evolution in basal fish lineages are variable. Additionally, removing these sequences resulted in a longer ancestral branch, and the difference among test results may in part be due to differences in statistical power. Regardless, these results provide strong and consistent evidence for increased dN/dS along the ancestral branch for GNAT2 but not for GNAT1.

Applying Zhang et al.'s branch-site models to the entire fish and tetrapods clades (rather than just to the ancestral branch connecting these two clades) did not yield evidence for clade-wide site-specific positive selection (P = 1.000 in all cases, for both data sets). However, parameter estimates did reveal patterns qualitatively consistent with our other analyses: (1) higher dN/dS in GNAT2 than in GNAT1; (2) higher dN/dS in the fish clade than in the tetrapod clade; (3) elevated dN/dS along the ancestral branch in the GNAT2 data set but not in the GNAT1 data set (see Supplementary Table 3 for more details).

Ancestral reconstructions and branch-site BEB analysis were used to gain further insight into early GNAT2 evolution. Twenty-five AA differences were identified between the inferred sequences at the bases of the tetrapod and ray-finned fish clades. The posterior support probabilities for the reconstructed sequences at these nodes were 0.972 and 0.977, respectively, when averaged over all 187 sites; unsurprisingly, support probabilities were lower for the 25 variable sites, ranging from 0.545 to 1.000 (mean = 0.894) at the ancestral tetrapod node and from 0.402 to 0.999 (mean = 0.848) at the ancestral fish node. Sixteen of these 25 variable AA sites were identified by branch-site BEB analysis as 'positively selected' along the branch uniting the tetrapod and fish clades with posterior probability >0.50 (Table 4). Two additional sites identified as positively
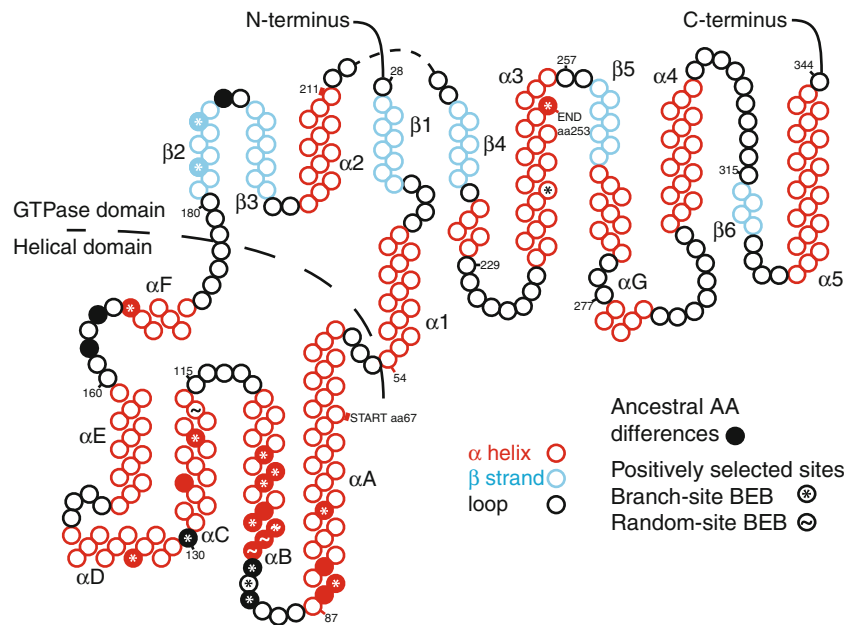
Fig. 4 GNAT snake plot showing the location of 'positively selected' sites identified by random-site and branch-site BEB analyses of the GNAT2 data set. *Filled circles* indicate AAs that differ between the ancestral sequences inferred for the tetrapod and ray-finned fish clades, tildes indicate random-site BEB sites, and asterisks indicate branch-site BEB sites. α Helices and β strands are color coded *red* and *blue*, respectively. The snake plot is drawn according to the GNAT1/GNAI crystal structure of Slep et al. (2001) (protein data bank ID 1FQJ). Site numbering follows that of bovine GNAT1, and the start and end points for the analyzed fragment are labeled (Color figure online)

Fig. 5 GNAT2 homology model (*orange*) aligned with the previously determined structures of GAP complex proteins RGS9 (*gray*) and GNB5 (*red*). Structures are shown in cartoon format. 'Positively selected' sites identified by random-site (*blue*) and branch-site (*green*) BEB analyses are shown in stick format, as are RGS9 and GNB5 sites found to be within approximately 5 Å of positively selected sites 101, 102, 104, and 117 by in silico protein structure alignment. Details on model construction are provided in the main text. Site numbering follows that of bovine GNAT1 (Color figure online)
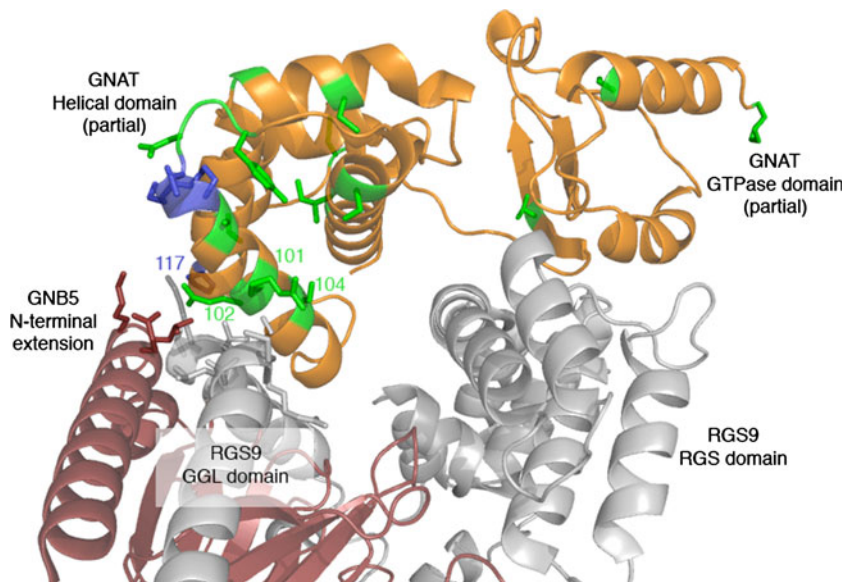
selected did not vary between the reconstructed AA sequences; these two discrepancies resulted from (1) the AA and codon analyses favouring different yet equally parsimonious reconstructions at position 92, and (2) the codon analysis highlighting multiple nonsynonymous changes that yielded no net AA change at position 242, which is fixed for serine, a sixfold degenerate AA. The spatial positions of the positively selected sites and the ancestral AA differences within the GNAT protein are displayed in Fig. 4. The majority of the identified sites were found to reside within the helical domain, and of those, most were found in the region spanning α helices A–C; this spatial clustering strongly suggests that selection is primarily targeting a function specific to the α helical domain. Visual inspection of

homology models constructed for the two reconstructed sequences revealed only minor differences involving slight displacement of individual amino acids rather than gross structural rearrangements (backbone root mean square deviation = 0.399 Å). Docking the homology model constructed using the tetrapod sequence onto a GAP complex crystal structure (following Cheever et al. 2008) revealed that several of the positively selected sites sterically clash with components of the GAP complex and that several other such sites are situated nearby (Fig. 5). Most prominently, positively selected sites 101, 102, and 104 were all found to be within 5 Å of a five amino acid stretch of the GGL domain (RSG9 residues 233–237) and three amino acids of the GNB5 N-terminal extension (GNB5 residues 29, 32, and 33). Interestingly, similar measurements revealed that BEB site 117, identified through random-site analysis of the GNAT2 data set, also resides within 5 Å of the RGS9 protein's GGL domain (sites 239, 241, and 242) (Fig. 5).

Discussion

Our molecular evolutionary analyses of vertebrate rod (GNAT1) and cone (GNAT2) transducin α subunits revealed interesting and distinct patterns of selection in the two data sets. Here, we summarize our findings in light of past studies of GNAT biochemistry and speculate on the possible adaptive relevance of cone transducin sequence variation in vertebrates.

Purifying selection has been the dominant force shaping the evolution of both GNAT1 and GNAT2 proteins; for both data sets and under all three-model frameworks (random-site, clade, and branch-site), dN/dS estimates are near zero for the majority of the data. This result is not surprising considering the key role vision plays for most vertebrates (Lythgoe 1979; Walls 1942), and the slow evolutionary rates seen for G protein α subunits in general (Yokoyama and Starmer 1992). However, for both genes, but especially for GNAT2, the degree of constraint varies nonrandomly across the gene; within the GTPase domain, the signature of purifying selection dominates, with most sites having dN/dS estimates near zero, but within the helical domain purifying selection is much weaker, with many sites having relatively high dN/dS estimates. While this pattern could simply indicate relaxed constraint in the helical domain, it seems likely that positive selection is playing a role, at least for GNAT2, as (1) site-specific dN/dS estimates are particularly high at clustered sites, and (2) branch-site analysis revealed the signature of strong positive selection early in vertebrate history. This pattern is consistent with different functional roles for the two GNAT domains. The GTPase domain is an evolutionarily ancient functional domain distantly related to single-domain small-

GTPase Ras proteins (Leipe et al. 2002; Lochrie et al. 1985a). Notably, regions of critical functional importance, such as the three 'switch' regions and sites implicated in GTP or PDE binding (Noel et al. 1993; Slep et al. 2001), are highly conserved and have dN/dS estimates of effectively zero, indicating strong purifying selection. Compared with the GTPase domain, the helical domain is a relatively recent evolutionary innovation, and its function is less well characterized (reviewed in Sprang 1997); proposed roles include modulating the GTPase domain's affinity for GTP or its catalytic activity or contributing to downstream effector binding. Our results suggest that one or more of those functions may have been targeted by positive selection at various points throughout GNAT2 history.

Several recent studies have highlighted a role for the helical domain in mediating GAP complex binding. Using chimeras of GNAT1 and GNAI, a homologous α subunit from the inhibitory G protein subfamily, Skiba et al. (1999) showed that the helical domain strongly influences the affinity of the RGS domain for the G protein α subunit. Moreover, Skiba et al. (1999) identified helix αB of the helical domain as a promising candidate for explaining this effect. Cheever et al. (2008) solved the 3D crystal structure of the GAP complex and reported that the helices αB and αC of the GNAT helical domain likely sterically clash with the GAP complex's RGS9 (specifically, the GGL domain) and GNB5 subunits. Finally, Soundararajan et al. (2008) examined the structural interaction between a diverse suite of isolated RGS domains and G protein α subunits and reported considerable heterogeneity in contact points; from this, they argued that the helical domain contacts play a key role in ensuring GAP specificity and potency. Given these findings, it is interesting to note that several sites within helices αA–αC of GNAT's helical domain were highlighted as positively selected and that some of these positively selected sites reside at Cheever et al.'s (2008) proposed GNAT–GAP complex interface; branch-site BEB sites 101, 102, and 104, and random-sites BEB site 117, reside within 5 Å of sites within the RGS9 protein's GGL domain and the GNB5 protein's N-terminal extension. The inferred amino acid substitutions at these sites are generally nonconservative; our ancestral reconstruction results indicate that two of the branch-site BEB sites underwent charge-changing substitutions along the ancestral GNAT2 branch (Q101R and K102Q), while the third underwent a polarity-changing substitution (N104S). Random-site BEB site 117, meanwhile, is represented by six AAs across the phylogeny (P, S, T, A, Q, and K); while some of the inferred substitutions at this site are conservative (e.g., S117T), others are decidedly not (e.g., P117Q). In vitro assays have confirmed that the GGL domain and the GNB5 N-terminal extension play an important role in ensuring

GAP complex specificity by its affinity for GNAT (Skiba et al. 2001). Our findings, which build on those of Cheever et al. (2008), indicate that this is likely due to specific interactions between key residues in the GGL domain and in GNB5 with key residues in GNAT helices αB and αC. Though it is currently difficult to speculate confidently on the direction of effect of specific substitutions, we predict that biochemical analysis of GNAT proteins with mutations at these sites will have altered GAP complex-mediated GTPase activity and, consequently, that photoreceptors harboring mutant GNAT proteins will have altered response recovery rates. Previous studies of rod and cone physiology have largely concluded that variation in photoreceptor deactivation kinetics reflects variation in GAP complex concentration (Burns and Pugh 2009, 2010); our findings suggest that variation at amino acid sites involved in the GNAT–GAP complex interaction likely plays an important role as well.

On this point, we note four caveats, some statistical and some biological. First, the identification of positively selected sites is an exercise in statistical inference, and the methods employed in this paper are not universally supported (Hughes 2007; Nozawa et al. 2009). However, simulation studies have validated the false-positive rates of these methods under a variety of conditions, and a number of studies have successfully used these methods to guide biochemical analyses (Yang and dos Reis 2011; Yang et al. 2009; Zhai et al. 2012). Likewise, the ancestral sequences are inferred, not known, and a full biochemical investigation of resurrected proteins should consider alternative ancestral sequences by sampling from the posterior distribution of ancestral states (Williams et al. 2006). Second, these methods cannot distinguish between positive selection that leads to novel biochemical phenotypes and positive selection that counteracts the effects of previously fixed deleterious mutations (i.e., compensatory evolution). Though such cases might not be considered 'adaptive evolution' in the typical sense, they would be indicative of regions or amino acid sites of functional importance. Third, the effects of individual substitutions may be subtle and difficult to detect using current *in vitro* methods. For example, Chen et al. (2010) found that replacing rod with cone transducin significantly affected response recovery in transgenic mouse rod photoreceptor cells, but Mao et al. (2013) did not, a difference that likely reflects variation in GNAT2 expression levels (GNAT2 levels were comparable to wild-type GNAT1 levels in the former study but less than 35 % of wild-type GNAT1 levels in the latter study). Fourth, *in vitro* GTPase assays that consider only the effect of the RGS domain will overlook important contributions made by the GGL domain, the GNB5 subunit, and the PDE γ subunit (Skiba et al. 2001). This may explain why a recent study by Gopalakrishna et al. (2012) reported no

significant difference in RGS domain-mediated GTPase activity between GNAT1/GNAI and GNAT2/GNAI proteins despite the two chimeric proteins possessing divergent helical domains (helical domain sequence identity = 66 %). Future studies that examine the sites highlighted by our analyses will have to take these points into consideration when designing and interpreting experiments.

The codon models we employed—probabilistic methods applied to multiple sequences united by a phylogeny—have their roots in approximate, pairwise methods that often fail to accurately estimate synonymous distances when sequences are highly divergent (Li et al. 1985; Miyata and Yasunaga 1980; Nei and Gojobori 1986; Perler et al. 1980; Yang and Nielsen 2000). Tetrapods and ray-finned fishes diverged approximately 425–430 Ma (Alfaro et al. 2009) and a large number of synonymous differences have accumulated between tetrapod and fish GNAT sequences over this time (sum of dS branch lengths between human and fugu = 3.4 for GNAT1 and = 2.9 for GNAT2 under the M0 codon model). Unfortunately, there are no concrete guidelines for establishing whether these data sets are too divergent to be analyzed reliably by modern codon models. We employed a simple rule-of-thumb, ensuring that branch-specific estimates of dS were always well below 2.0 synonymous substitutions per synonymous site. Alternatively, simulation studies can provide useful insights into the limits of codon models. Although we did not conduct such analyses here, two recent studies (Gharib and Robinson-Rechavi 2013; Studer et al. 2008) are particularly relevant to our analyses, as both conditioned their simulations on preliminary analyses of vertebrate-wide data sets and both applied the branch-site test to the branch connecting the tetrapod and fish clades. These studies, in brief, demonstrated that the branch-site test is conservative and that the test remains so even when branch lengths are artificially lengthened, when GC-content is varied either across the alignment or across the phylogeny, or when positive selection occurs elsewhere in the phylogeny (in violation of branch-site model assumptions). Studer et al. (2008) highlighted the role of dense phylogenetic sampling in obtaining reliable test results, and it is notable that these simulation studies used data sets comprised of fewer sequences than present in our GNAT data sets. Of course, these simulation results were obtained using error-free alignments, and it is known that alignment errors can results in false signatures of positive selection (Fletcher and Yang 2010; Jordan and Goldman 2012); fortunately, this seems unlikely to affect our analyses due to the highly conserved primary and tertiary structures of G protein α subunits. Together, these findings suggest that our GNAT data sets should provide no major problems for codon model analyses in spite of the considerable

evolutionary distances involved. We note here that codon models have been used to study even more ancient evolutionary events than the ones we considered here [e.g., gene duplications that predate the vertebrate radiation: (Carretero-Paulet et al. 2013; Weadick and Chang 2009)]. Extending the work of Studer et al. (2008) and Gharib and Robinson-Rechavi (2013) to consider such data sets will provide useful insights into the effective limits of codon model analyses. In this context, it will be useful to explicitly consider the effects of protein structure so that more reliably analyzed data sets—such as those based on proteins with highly conserved tertiary structures—can be distinguished from those likely to produce misleading results.

In addition to the dramatic increase in dN/dS reported by branch-site analysis of the GNAT2 data set, our clade model analysis indicates that patterns of sequence evolution varied between the fish and tetrapod clades, with dN/dS lower in the tetrapod clade. A similar pattern of phylogenetic variation is present for GNAT1, as well, even though our GNAT1 analyses failed to provide convincing evidence for positive selection. The fact that many fish species have well-developed visual systems and complex visual ecologies (Lythgoe 1979; Walls 1942) argues against a simple 'relaxed constraint' explanation for this difference. Alternatively, the severity of the trade-off between absolute sensitivity and temporal resolution may be heightened in the underwater environment due to lower light levels; this could impose diversifying selection on GNAT2 sequences as different species balance the trade off in different ways, given each species' visual niche (i.e., depth, activity patterns). Physiological differences may play a role as well, as the tetrapods that comprise our data sets are primarily homeotherms, and in these species, phototransduction will occur under fairly consistent thermal conditions. Fish, by contrast, are generally poikilotherms, and temperature can have large effects on photoreceptor response and recovery (Fritsches et al. 2005; Lamb 1984). By demonstrating variation in GNAT selective constraint between tetrapods and fishes, our results provide a foundation for further study of ecologically driven GNAT evolution in vertebrates. Targeted examination of relatively closely related species that have repeatedly invaded different niches may be particularly informative, as has been shown for the rod opsins of deep and shallow water cichlids (Nagai et al. 2011). At larger phylogenetic scales, the inclusion of GNAT sequences from key taxa such as lampreys and coelacanths (Lagman et al. 2012; Muradov et al. 2008) may yield interesting insights into key transition in vertebrate history, such as the origin of rod photopigments (Trezise and Collin 2005) and the transition from water to land by early tetrapods (Long and Gordon 2004). Such studies will further enhance our understanding of the importance of sequence-level variation in GNAT proteins. Furthermore, by employing a comparative approach in order to identify particularly unusual or divergent sequences, such efforts may help direct future biochemical analyses of GNAT structure and function.

References

Abascal F, Zardoya R, Posada D (2005) Prottest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc Natl Acad Sci 106:13410–13414

Anisimova M, Liberles DA (2012) Detecting and understanding natural selection. In: Cannarozzi GM, Schneider A (eds) Codon evolution: mechanisms and models. Oxford University Press, Oxford, pp 73–96

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Bielawski JP, Yang ZH (2004) A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. J Mol Evol 59:121–132

Burns ME, Pugh EN (2009) Rgs9 concentration matters in rod phototransduction. Biophys J 97:1538–1547

Burns ME, Pugh EN (2010) Lessons from photoreceptors: turning off g-protein signaling in living cells. Physiology 25:72–84

Carretero-Paulet L, Albert VA, Fares MA (2013) Molecular evolutionary mechanisms driving functional diversification of the hsp90a family of heat shock proteins in eukaryotes. Mol Biol Evol 30:2035–2043

Cheever ML, Snyder JT, Gershburg S, Siderovski DP, Harden TK, Sondek J (2008) Crystal structure of the multifunctional g beta 5-rgs9 complex. Nat Struct Mol Biol 15:155–162

Chen CK, Woodruff ML, Chen FS, Shim H, Cilluffo MC, Fain GL (2010) Replacing the rod with the cone transducin alpha subunit decreases sensitivity and accelerates response decay. J Physiol (London) 588:3231–3241

Cowan CW, Fariss RN, Sokal I, Palczewski K, Wensel TG (1998) High expression levels in cones of rgs9, the predominant gtpase accelerating protein of rods. Proc Natl Acad Sci USA 95:5351–5356

Delport W, Poon AFY, Frost SDW, Pond SLK (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457

Ebrey T, Koutalos Y (2001) Vertebrate photoreceptors. Prog Retin Eye Res 20:49–94

Fain GL, Hardie R, Laughlin SB (2010) Phototransduction and the evolution of photoreceptors. Curr Biol 20:R114–R124

Fletcher W, Yang Z (2010) The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. Mol Biol Evol 27:2257–2267

Fritsches KA, Brill RW, Warrant EJ (2005) Warm eyes provide superior vision in swordfishes. Curr Biol 15:55–58

Gharib WH, Robinson-Rechavi M (2013) The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in gc. Mol Biol Evol 30:1675–1686

Gojobori T (1983) Codon substitution in evolution and the "saturation" of synonymous changes. Genetics 105:1011–1027

Goldman N, Whelan S (2000) Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Mol Biol Evol 17:975–978

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

Gopalakrishna KN, Boyd KK, Artemyev NO (2012) Comparative analysis of cone and rod transducins using chimeric galpha subunits. Biochemistry 51:1617–1624

He W, Cowan CW, Wensel TG (1998) Rgs9, a gtpase accelerator for phototransduction. Neuron 20:95–102

Hu G, Wensel TG (2002) R9ap, a membrane anchor for the photoreceptor gtpase accelerating protein, rgs9-1. Proc Natl Acad Sci USA 99:9755–9760

Huelsenbeck JP, Rannala B (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science 276:227–232

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity (Edinb) 99:364–373

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Jordan G, Goldman N (2012) The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol Biol Evol 29:1125–1139

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol 28:3033–3043

Lagman D, Sundstrom G, Ocampo Daza D, Abalo XM, Larhammar D (2012) Expansion of transducin subunit gene families in early vertebrate tetraploidizations. Genomics 100:203–211

Lamb TD (1984) Effects of temperature changes on toad rod photocurrents. J Physiol 346:557–578

Lambright DG, Noel JP, Hamm HE, Sigler PB (1994) Structural determinants for activation of the alpha-subunit of a heterotrimeric g-protein. Nature 369:621–628

Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

Leipe DD, Wolf YI, Koonin EV, Aravind L (2002) Classification and evolution of p-loop gtpases and related atpases. J Mol Biol 317:41–72

Lerea CL, Somers DE, Hurley JB, Klock IB, Buntmilam AH (1986) Identification of specific transducin alpha-subunits in retinal rod and cone photoreceptors. Science 234:77–80

Li W-H, Wu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Lochrie MA, Hurley JB, Simon MI (1985a) Sequence of the alpha subunit of photoreceptor g protein: homologies between transducin, ras, and elongation factors. Science 228:96–99

Lochrie MA, Hurley JB, Simon MI (1985b) Sequence of the alpha-subunit of photoreceptor-g protein—homologies between transducin, ras, and elongation-factors. Science 228:96–99

Long JA, Gordon MS (2004) The greatest step in vertebrate history: a paleobiological review of the fish-tetrapod transition. Physiol Biochem Zool 77:700–719

Loytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci USA 102:10557–10562

Loytynoja A, Vilella AJ, Goldman N (2012) Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. Bioinformatics 28:1684–1691

Lythgoe JN (1979) The ecology of vision. Clarendon Press, Oxford

Lyubarsky A, Chen CK, Naarendorp F, Zhang X, Wensel T, Simon M, Pugh E (2001) Rgs9-1 is required for normal inactivation of mouse cone phototransduction. Mol Vis 7:71–78

Maddison DR, Schulz K-S (eds) (2007) The tree of life web project. http://tolweb.org

Makino ER, Handy JW, Li TS, Arshavsky VY (1999) The gtpase activating factor for transducin in rod photoreceptors is the complex between rgs9 and type 5 g protein beta subunit. Proc Natl Acad Sci USA 96:1947–1952

Mao W, Miyagishima KJ, Yao Y, Soreghan B, Sampath AP, Chen J (2013) Functional comparison of rod and cone galphat on the regulation of light sensitivity. J Biol Chem 288:5257–5267

Maynard Smith J (1994) Estimating selection by comparing synonymous and substitutional changes. J Mol Evol 39:123–128

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–36

Muradov H, Kerov V, Boyd KK, Artemyev NO (2008) Unique transducins expressed in long and short photoreceptors of lamprey petromyzon marinus. Vis Res 48:2302–2308

Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11:715–724

Nagai H, Terai Y, Sugawara T, Imai H, Nishihara H, Hori M, Okada N (2011) Reverse evolution in rh1 for adaptation of cichlids to water depth in Lake Tanganyika. Mol Biol Evol 28:1769–1776

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

Noel JP, Hamm HE, Sigler PB (1993) The 2.2-angstrom crystal-structure of transducin-alpha complexed with gtp-gamma-s. Nature 366:654–663

Nordstrom K, Larsson TA, Larhammar D (2004) Extensive duplications of phototransduction genes in early vertebrate evolution correlate with block (chromosome) duplications. Genomics 83:852–872

Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. Proc Natl Acad Sci USA 106:6700–6705

Oldham WM, Hamm HE (2008) Heterotrimeric g protein activation by g-protein-coupled receptors. Nat Rev Mol Cell Bio 9:60–71

Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T (2010a) Guidance: a web server for assessing alignment confidence scores. Nucleic Acids Res 38:W23–W28

Penn O, Privman E, Landan G, Graur D, Pupko T (2010b) An alignment confidence score capturing robustness to guide tree uncertainty. Mol Biol Evol 27:1759–1767

Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolodner R, Dodgson J (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20:555–566

Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234:779–815

Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524

Skiba NP, Yang CS, Huang T, Bae H, Hamm HE (1999) The alpha-helical domain of galphat determines specific interaction with regulator of g protein signaling 9. J Biol Chem 274:8770–8778

Skiba NP, Martemyanov KA, Elfenbein A, Hopp JA, Bohm A, Simonds WF, Arshavsky VY (2001) Rgs9-g beta 5 substrate selectivity in photoreceptors. Opposing effects of constituent domains yield high affinity of rgs interaction with the g protein-effector complex. J Biol Chem 276:37365–37372

Slep KC, Kercher MA, He W, Cowan CW, Wensel TG, Sigler PB (2001) Structural determinants for regulation of

phosphodiesterase by a g protein at 2.0 angstrom. Nature 409: 1071–1077

Sondek J, Lambright DG, Noel JP, Hamm HE, Sigler PB (1994) Gtpase mechanism of g proteins from the 1.7-angstrom crystal-structure of transducin alpha-center-dot-gdp-center-dot-alf4(−). Nature 372:276–279

Soundararajan M, Willard FS, Kimple AJ, Turnbull AP, Ball LJ, Schoch GA, Gileadi C, Fedorov OY, Dowler EF, Higman VA, Hutsell SQ, Sundstrom M, Doyle DA, Siderovski DP (2008) Structural diversity in the rgs domain and its interaction with heterotrimeric g protein alpha-subunits. Proc Natl Acad Sci USA 105:6457–6462

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

Studer RA, Penel S, Duret L, Robinson-Rechavi M (2008) Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. Genome Res 18:1393–1402

Swanson WJ, Nielsen R, Yang QF (2003) Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20

Tamura K, Dudley J, Nei M, Kumar S (2007) Mega4: molecular evolutionary genetics analysis (mega) software version 4.0. Mol Biol Evol 24:1596–1599

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

Trezise AE, Collin SP (2005) Opsins: evolution in waiting. Curr Biol 15:R794–R796

Walls GL (1942) The vertebrate eye and its adaptive radiation. Oxford, England: Cranbrook Institute of Science, p 785

Weadick CJ, Chang BS (2009) Molecular evolution of the œ ≤ œ - ≥ lens crystallin superfamily: evidence for a retained ancestral function in œ ≥ n crystallins? Mol Biol Evol 26:1127–1142

Weadick CJ, Chang BSW (2012) An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. Mol Biol Evol 29:1297–1300

Wiederstein M, Sippl MJ (2007) Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410

Williams PD, Pollock DD, Blackburne BP, Goldstein RA (2006) Assessing the accuracy of ancestral protein reconstruction methods. PLoS Comput Biol 2:e69

Wittinghofer A (1994) The structure of transducin-g(alpha-t)—more to view than just ras. Cell 76:201–204

Yang ZH (1994) Maximum-likelihood phylogenetic estimation from DNA-sequences with variable rates over sites—approximate methods. J Mol Evol 39:306–314

Yang ZH (2007) Paml 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, dos Reis M (2011) Statistical properties of the branch-site test of positive selection. Mol Biol Evol 28:1217–1228

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yang ZH, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Yang ZH, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino-acid-sequences. Genetics 141:1641–1650

Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Yang Z, Nielsen R, Goldman N (2009) In defense of statistical methods for detecting positive selection. Proc Natl Acad Sci USA 106:E95; author reply E96

Yau KW, Hardie RC (2009) Phototransduction motifs and variations. Cell 139:246–264

Yokoyama S, Starmer WT (1992) Phylogeny and evolutionary rates of g protein alpha subunit genes. J Mol Evol 35:230–238

Yoshida I, Sugiura W, Shibata J, Ren FR, Yang ZH, Tanaka H (2011) Change of positive selection pressure on hiv-1 envelope gene inferred by early and recent samples. PLoS ONE 6(4):e18630

Zhai W, Nielsen R, Goldman N, Yang Z (2012) Looking for Darwin in genomic sequences—validity and success of statistical methods. Mol Biol Evol 29:2889–2893

Zhang X, Wensel TG, Kraft TW (2003) Gtpase regulators and photoresponses in cones of the eastern chipmunk. J Neurosci 23:1287–1297

Zhang JZ, Nielsen R, Yang ZH (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479