Molecular Evolution of the $\beta\gamma$ Lens Crystallin Superfamily: Evidence for a Retained Ancestral Function in γN Crystallins?

Cameron J. Weadick and Belinda S.W. Chang*†*

*Department of Ecology and Evolution, University of Toronto, Toronto, Ontario, Canada; and †Department of Cell and Systems Biology, and Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, Canada

Within the vertebrate eye, $\beta\gamma$ crystallins are extremely stable lens proteins that are uniquely adapted to increase refractory power while maintaining transparency. Unlike α crystallins, which are well-characterized, multifunctional proteins that have important functions both in and out of the lens, $\beta\gamma$ lens crystallins are a diverse group of proteins with no clear ancestral or contemporary nonlens role. We carried out phylogenetic and molecular evolutionary analyses of the $\beta\gamma$-crystallin superfamily in order to study the evolutionary history of the γN crystallins, a recently discovered, biochemically atypical family suggested to possess a divergent or ancestral function. By including nonlens, $\beta\gamma$-motif-containing sequences in our analysis as outgroups, we confirmed the phylogenetic position of the γN family as sister to other γ crystallins. Using maximum likelihood codon models to estimate lineage-specific nonsynonymous-to-synonymous rate ratios revealed strong positive selection in all of the early lineages within the $\beta\gamma$ family, with the striking exception of the lineage leading to the γN crystallins which was characterized by strong purifying selection. Branch-site analysis, used to identify candidate sites involved in functional divergence between γN crystallins and its sister clade containing all other γ crystallins, identified several positively selected changes at sites of known functional importance in the $\beta\gamma$ crystallin protein structure. Further analyses of a fish-specific γN crystallin gene duplication revealed a more recent episode of positive selection in only one of the two descendant lineages (γN2). Finally, from the guppy, *Poecilia reticulata*, we isolated complete γN1 and γN2 coding sequence data from cDNA and partial coding sequence data from genomic DNA in order to confirm the presence of a novel γN2 intron, discovered through data mining of two pufferfish genomes. We conclude that the function of the γN family likely resembles the ancestral vertebrate $\beta\gamma$ crystallin more than other $\beta\gamma$ families. Furthermore, owing to the presence of an additional intron in some fish γN2 crystallins, and the inferred action of positive selection following the fish-specific γN duplication, we suggest that further study of fish γN crystallins will be critical in further elucidating possible ancestral functions of γN crystallins and any nonstructural role they may have.

Introduction

The evolution of the vertebrate eye has long fascinated and confounded biologists (Darwin 1859; Land and Fernald 1992; Fernald 2006), and past work on lens crystallins, a diverse and heterogeneous group of structural proteins that contribute to the transparency and refractory power of the lens, has revealed that at least some components of the eye were recruited from ancestral roles unrelated to photosensation (Wistow 1995; Piatigorsky 2007). For example, α crystallins, a major component of all vertebrate lenses, are members of a larger gene family that includes small heat-shock proteins (Franck et al. 2004), with some members still retaining the ability to act as molecular chaperones (Piatigorsky 2007). Similarly, several taxon-specific crystallins, found only in specific vertebrate lineages, are closely related to various metabolic enzymes, if not the exact same protein functioning in two entirely different roles (Wistow 1993). The same gene that encodes ϵ-crystallin, a crystallin found in the lenses of birds and crocodilians, also codes for the metabolic enzyme lactate dehydrogenase. Crystallins are generally found to be extremely stable (necessary because lens cells enucleate during early development and subsequently lack the translational machinery to generate new proteins), highly soluble (necessary as high protein concentration increases the refractory power of the lens), and to display short-range ordering (the ability to exist at high concentrations without scattering light and causing lens opacity) (reviewed by Bloemendal et al. 2004). Whether or not these proteins were initially selected for use in a lens role because of their ancestral functions, or simply because they possessed properties ideal for use in a transparent and refractive lens, remains as yet unknown, but it has been suggested that the particular cellular conditions present during lens development favored the co-option of stress-tolerant and/or stress-responsive proteins such as α crystallins for use in the ancestral proto-lens (de Jong et al. 1989; Wistow et al. 1995). Such protein multifunctionality may introduce constraints that prevent selection from simultaneously optimizing the protein's ability to carry out both its crystallin and noncrystallin roles, leading to adaptive conflict (Wistow 1993; Hughes 1994). One route through which this conflict could be resolved would be through gene duplication and subsequent specialization (e.g., Des Marais and Rausher 2008).

It has similarly been suggested that $\beta\gamma$ crystallins, another group of broadly distributed lens crystallins, were recruited for use in the lens because of a connection to stress response (de Jong et al. 1989; Wistow et al. 1995). Members of the $\beta\gamma$ lens crystallin superfamily include the oligomeric βA and βB, as well as the monomeric γS, γTerrestrial, and γAquatic (or γM) families. Structurally, $\beta\gamma$ crystallins are rather unusual in that they are composed of double Greek key motifs, β-strand motifs that intercalate to form a highly symmetrical β-sandwich domain, though usually only one tyrosine corner is present per domain (reviewed by Bloemendal et al. 2004). γ Crystallin monomers are comprised of two β-sandwich domains arranged, again, with a high degree of symmetry. This is in contrast to the oligomeric β crystallins, in which the two β-sandwich domains form intermolecular pairings. Striking features of $\beta\gamma$ crystallins in general include these exchanges about various

Key words: $\beta\gamma$ crystallins, γN crystallins, gene duplication, protein structure function, novel intron, calcium binding.

E-mail: belinda.chang@utoronto.ca.

Mol. Biol. Evol. 26(5):1127–1142. 2009
doi:10.1093/molbev/msp028
Advance Access publication February 20, 2009

axes of symmetry, another example occurring between the third β-strand of a Greek key motif with its partner; these unique features may be important in contributing to the high intrinsic stability characteristic of these proteins. Related proteins with similarly stable properties have also been found in bacteria and single-celled eukaryotes, where they are thought to play a role in stress-induced sporulation (Wistow et al. 1985). However, there is considerable divergence between these sporulation-related proteins and βγ lens crystallins. More closely related proteins have been documented (reviewed by Piatigorsky 2007), including several epithelial-specific proteins (EP) of the newt *Cynops pyrrhogaster* (Takabatake et al. 1992; Wistow et al. 1995; Ogawa et al. 1997, 1998), a βγ domain containing protein (AIM1) whose upregulation appears to have antitumorigenic properties in melanoma cell lines (Ray et al. 1997), and a single-domain βγ crystallin found in the palps and otolith of the larval urochordate *Ciona intestinalis* (Shimeld et al. 2005). Unfortunately, the specific biochemical role played by the βγ domains in these various proteins is unknown and determining the ancestral nonlens role of βγ crystallins from these examples is difficult. The early history of the βγ lens crystallins is therefore unclear.

Clues to the ancestral function of βγ lens crystallins may lie within a recently discovered family, the γN crystallins (Wistow et al. 2005). γN crystallins, represented by a single copy in several tetrapods and two copies in zebrafish (*Danio rerio*), display greater sequence similarity with β and γ lens crystallins than any other known βγ domain–containing protein. Though phylogenetic analysis was unable to fully resolve the position of the γN family in the larger βγ superfamily, analysis of gene structure and solution behavior suggested that this new family nests sister to other γ crystallins (Wistow et al. 2005). Members of the γN family were found to possess a hybrid exon–intron gene structure, with the two motifs of the C-terminal domain being separated by an intron, like the β crystallins, but with the entire N-terminal domain being coded by a single exon, as in the γ crystallins. Additionally, this new family was found to exist as a monomer in solution, like other γ crystallins. Quite unexpectedly, and despite its close evolutionary proximity to other βγ lens crystallins, mouse (*Mus musculus*) γN was found to be both highly unsoluble and unstable (Wistow et al. 2005). The evolutionary and mechanistic reasons why mouse γN possesses such unusual biochemical properties are currently unknown, but as high stability and solubility are hallmarks of most other crystallins (Bloemendal et al. 2004), Wistow et al. (2005) suggested that the γN crystallins might possess a distinct function within the βγ crystallin superfamily, possibly even an ancestral, prelens role.

Recent years have seen widespread application of tree-based, d_N/d_S estimation techniques (Yang and Bielawski 2000) to many different gene families, where they have proven to be useful methods for investigating and generating hypotheses with respect to the evolution of protein structure and function (e.g., Yang and Nielsen 1998; Zhang et al. 1998; Merritt and Quattro 2001; Swanson et al. 2003; Sawyer et al. 2005; Sassi et al. 2007; Sweeney et al. 2007; Weadick and Chang 2007; Des Marais and Rausher 2008). However, these methods have been largely unexplored for the vertebrate lens crystallins (though see Sweeney et al. 2007 for an example involving squid lens crystallins). In this paper, we use these methods to explore the molecular evolution of the βγ crystallin superfamily, with particular emphasis on gaining insights into why, mechanistically, γN crystallins display such divergent biochemical properties (Wistow et al. 2005). We also obtained additional fish γN sequences through a combination of laboratory sequencing and genome database mining and investigated whether divergent selection might have been associated with a fish-specific duplication of γN crystallins. Finally, we clarify patterns of intron evolution in the βγ superfamily and provide evidence for the presence of a novel intron in the γN2 crystallin of some fish species.

Materials and Methods

Cloning and Sequencing of Guppy γN Crystallins and Novel γN2 Intron

The complete coding sequences of two γN crystallins from the guppy, *P. reticulata*, were obtained by screening, cloning, and sequencing from a cDNA library. This library was derived from head tissue from a mature, male guppy and created using the SMART cDNA Library Construction Kit (BD Biosciences, San Jose, CA), following RNA extraction using an RNeasy MiniKit (Qiagen, Germantown, MD). Our amplification strategy involved degenerate-primer polymerase chain reaction (PCR), followed by 5′ and 3′ rapid amplification of CDNA ends PCR (Frohman et al. 1988), and finally the amplification of complete coding sequences using specific 5′ and 3′ untranslated-region primers (Pret_γN1_F_5UT 5′-gtttggaggactttaacagtgg; Pret_γN1_R_3UT 5′-gatatatcctgaaagcgtggca; Pret_γN2_F_5UT 5′-cactcagtgtgtggttgaggga; and Pret_γN2_R_3UT 5′-tactgatctgcctgctgagga) using FastStart Taq polymerase (Roche, Basel, Switzerland). Purified PCR products were ligated into TA-overhang plasmid vectors (Invitrogen, Carlsbad, CA), cultured in *Escherichia coli*, and sequenced on an ABI-377 capillary sequencer (Applied Biosystems, Foster City, CA) using vector primers. In order to check for a possible extra intron in the guppy γN2 crystallin (described below), we also used specific primers designed according to the cDNA sequence data to amplify from guppy genomic DNA (Pret_γN1_aa18_F 5′-ggcaggaagctggaactctgcagc; Pret_γN1_aa66_R 5′-gctccagaatgtactgctggccct; Pret_γN2_aa16_F 5′-ttcaccggcaggaagctggtgg; and Pret_γN2_aa88_R 5′-tgggcttgcaggatcccatgtgg). Genomic DNA was isolated using a DNeasy Blood and Tissue Kit (Qiagen). All other steps involved in obtaining sequence data from genomic DNA were identical to those used with the cDNA-derived data.

Database Searches

Owing to its relatively recent discovery (Wistow et al. 2005), few annotated and predicted γN sequences were available on public databases during the assembly of our data sets. We therefore used BLAST searches to mine Genbank and Ensembl for unannotated γN crystallin sequences, including the clawed frog *Xenopus tropicalis*, the rhesus macaque *Macaca mulatta*, the short-tailed opossum

Monodelphis domestica, and two pufferfish, the green spotted puffer *Tetraodon nigroviridis* and fugu puffer *Takifugu rubripes*. In most cases, we only found a single region of high similarity to a γN crystallin query, but two regions were found in both pufferfish genomes, similar to the zebrafish genome. In both of the pufferfish genomes, one of the two copies (the γN2 paralog) was found to possess a novel intron in the first domain (discussed below). In order to clarify the phylogenetic position of γN crystallins within the larger βγ crystallin superfamily, a variety of βγ lens crystallin coding nucleotide sequences were obtained from Genbank using a combination of keyword and BLAST searches. Sequences from all major vertebrate groups for which extensive surveys of βγ crystallins have occurred were included (mammals, amphibians, reptiles, birds, and bony fishes). Throughout this paper, amino acid numbering will follow that of the methionine-cleaved cow (*Bos taurus*) γB crystallin (PDB 1AMM; Kumaraswamy et al. 1996). Finally, in addition to the various γ (γS, γTerrestrial, and γAquatic [or γM] crystallins, which we group as γSTA crystallins), β (βA and βB), and recently discovered γN crystallins, βγ-motif containing epithelial proteins (EPs) from the Japanese firebelly newt (*C. pyrrhogaster*) were included as outgroups (Takabatake et al. 1992; Ogawa et al. 1997, 1998).

Alignment

We aligned amino acid translated crystallin sequences using ClustalW (Thompson et al. 1994) as implemented in MEGA3 (Kumar et al. 2004). Alignments were attempted using the BLOSSUM, PAM, IDENTITY, and GONNET residue weight matrices, and the resulting alignments were compared using the HoT alignment evaluation method (Landan and Graur 2007). This method, which evaluates the relative consistency of the alignment method and the relative accuracy of the resulting alignment (Hall 2008), led us to select the alignment created using the GONNET substitution weight matrix for further investigation. The N- and C-terminal tails of some sequences, which are highly variable in both length and sequence, were removed prior to alignment. All internal alignment sites for which fewer than five ingroup sequences contained data were excluded from the phylogenetic and molecular evolutionary analyses (βγ-alignment). Further molecular evolutionary analyses were carried out on a reduced version of this data set containing only γSTA and γN crystallins (γ-alignment). Finally, phylogenetic and molecular evolutionary analyses were also carried out on a second alignment consisting of only γN crystallin sequences (γN-alignment); because the γN sequences contain no indels, they were aligned by eye.

Phylogenetic Analysis

We estimated the phylogeny of the vertebrate βγ lens crystallin superfamily (βγ-alignment) as well as the γN family (γN-alignment) using a Bayesian approach (Holder and Lewis 2003), as implemented in MrBayes 3 (Ronquist and Huelsenbeck 2003). Analyses were carried out on both nucleotide and translated amino acid data sets. As proper model choice is critical for Bayesian and other model-based approaches to phylogeny estimation (Posada and Crandall 2001), we employed various models of substitution and/or partitioning strategies and used Bayes factors (BFs) to compare the results (Kass and Raftery 1995). BFs, calculated as the ratio of marginal likelihood scores for competing models, are typically interpreted according to the guidelines of Kass and Raftery (1995); values of 2*ln(BF) between 2 and 5 constitute "positive" evidence, between 5 and 10 "strong" evidence, and over 10 "decisive" evidence, with negative values indicating support for the competing model. The marginal likelihood scores were estimated in MrBayes as the harmonic mean of sampled likelihood scores. Bayesian analysis provides posterior probability (PP) values as measures of support for particular topological divisions. In addition to PP values, BFs were also used to evaluate the amount of evidence in favor of particular clades; this was done by constraining groups of sequences to be monophyletic and comparing the constrained and unconstrained marginal likelihood scores. It should be noted, however, that the reliability of the suggested BF cutoff values when used in phylogenetic analysis has been evaluated more thoroughly for model choice and alignment partitioning decisions (Nylander et al. 2004; Brown and Lemmon 2007) than it has for topology investigation (Danforth et al. 2006; Dohrmann et al. 2006).

For the nucleotide version of the βγ superfamily data set, we attempted both standard and partitioned Bayesian analyses (Nylander et al. 2004; Brown and Lemmon 2007), with separate substitution models being employed for each codon position. For the standard Bayesian analysis, model selection, carried out using hierarchical likelihood ratio tests (LRTs; Felsenstein 1981) as implemented in MrModeltest (Posada and Crandall 2001; Nylander 2004), led us to use a general time reversible (GTR; Tavare 1986) nucleotide substitution model, with a proportion of sites classified as invariant (+I; Hasegawa et al. 1985) and a Γ distribution (+Γ; Yang 1994) used to describe among-site rate variation. For the partitioned Bayesian analysis, we used a GTR + I + Γ model for the first codon position and separate GTR + Γ models for the second and third codon positions. For the translated amino acid version of the data set, we estimated the phylogeny using a Jones, Taylor, and Thornton (JTT) model (Jones et al. 1992) of amino acid substitution, both with and without a correction for among-site rate variation (+Γ). Two independent Bayesian Markov chain Monte Carlo runs, each including three heated and one cold chain, were performed for all analyses, sampling every 100 generations. Nucleotide analyses were carried out for 10 million generations, whereas for amino acid analyses, an arbitrarily large number of generations were used in combination with a stop rule that stopped the run when the standard deviation of split frequencies fell to 0.01. In all cases, the first 25% of posterior samples were considered burn-in and discarded; visual inspection of the sampled likelihood scores and the potential scale reduction factors values suggested that this was sufficient to ensure convergence. A BF comparison of the standard and partitioned nucleotide approaches yielded 2*ln(BF) = 1,196.12, which constitutes decisive evidence against the

standard Bayesian analysis (Kass and Raftery 1995). Similarly, a BF comparison was used to compare the two amino acid models, and in this case, 2*ln(BF) = 450.32, indicating decisive evidence in favor of the more complex JTT + Γ model. As such, only the partitioned nucleotide and JTT + Γ amino acid analyses are discussed in the results.

We adopted a similar approach to estimating the γN family phylogeny as we did for the larger $\beta\gamma$ superfamily. The only exception was that we did not employ a stop rule for the amino acid analyses but instead ran the amino acid analyses for 5 million generations. LRTs among nucleotide models led us to select a symmetrical (SYM; Zharkikh 1994) substitution model with a Γ distribution (+Γ) for the nonpartitioned analysis, and SYM + Γ, GTR + Γ, and GTR + Γ for the three codon positions in the partitioned analysis, respectively. The partitioned analysis was favored over the standard analysis according to a BF comparison, with 2*ln(BF) = 527.06 (decisive evidence). Amino acid analysis was carried out using JTT and JTT + Γ models, as above, and again the JTT + Γ model was favored over the JTT model, with 2*ln(BF) = 9.44 (strong evidence). Because of these BF results, we only discuss the partitioned nucleotide and JTT + Γ amino acid analyses in the results.

Molecular Evolutionary Analysis

In order to estimate the form and strength of selection operating on the $\beta\gamma$ crystallin gene family, we used maximum likelihood methods to estimate d_N/d_S (or ω), the ratio of the rates of nonsynonymous (d_N) and synonymous (d_S) nucleotide substitutions, by employing a variety of codon substitution models as implemented in the codeml program of the PAML 3.15 software package (Yang 1997). Assuming no selection pressure, coding sequences will evolve neutrally, and nonsynonymous and synonymous substitutions should accumulate at equal rates, resulting in a d_N/d_S value equal to one (Kimura 1977; Yang and Bielawski 2000). Positive (or diversifying) selection is indicated by d_N/d_S values greater than one, whereas negative (or purifying) selection is indicated by d_N/d_S values near zero. We fit a variety of nested random-sites, branch, and branch-sites models to the $\beta\gamma$-, γ-, and γN-alignments in order to test for variation in selection pressure and/or positive selection. Likelihood differences between nested models were evaluated using LRTs. All PAML analyses were carried out using the F3×4 codon frequency estimation method and, where possible, models were run multiple times from different starting d_N/d_S values (below, at, and above $d_N/d_S = 1$) in order to check for convergence; only the results for runs with the best likelihood scores were considered further.

To test for pervasive among-site variation in d_N/d_S in the $\beta\gamma$- and γN-alignments, we fit the following random-sites models using PAML: M0, M1a, M2a, M3, M7, and M8 (Yang et al. 2000, 2005). Comparing model M3 against the null M0 tests for variation in long-term selection pressure among alignment sites, while comparing M2a and M8 against M1a and M7, respectively, tests for long-term positive selection acting on particular alignment sites. We also compared the M8 model with the M8a null model, as this comparison is believed to better fit the assumptions of LRTs and may provide a more accurate test of pervasive positive selection under some circumstances (Swanson et al. 2003). To investigate how the strength and form of selection operating on the early history of the γN family compared with selection pressure in other ancient branches of the phylogeny, we fit a branch model (Yang 1998; Yang and Nielsen 1998), which we term the 6fgb (6 foreground branches) model, to the $\beta\gamma$-alignment where the branch immediately ancestral to the γN family, as well as five other comparably ancient branches, were identified as foreground branches with each having independently estimated d_N/d_S values. Specifically, the six foreground branches were those leading to the γN, γSTA, βA, βB, all γ, and all β crystallins (fig. 1*a*). The 6fgb model was compared against the basic M0 model, which estimates a single d_N/d_S for the entire data set.

To identify sites likely involved in functional divergence between the γN and γSTA crystallins, we used a branch-site test (Test 2 of Yang and Nielsen 2002; Zhang et al. 2005) on a reduced data set including only these two groups (γ-alignment). Applying the branch-site model to the γ-alignment, with the branch connecting the γN and γSTA clades defined as foreground, allowed us to use Bayes empirical Bayes (BEB; Yang et al. 2005) analysis to identify positively selected codons that may have contributed to functional differentiation between the two groups. We also used DIVERGE 1.04 (Gu and Vander Velden 2002) to identify amino acid sites experiencing different degrees of constraint within the γN and γSTA crystallin clades; this method also employs a maximum likelihood approach and uses a Bayesian framework to identify specific sites that show a signature of functional divergence (Gu 1999). Sites identified as the likely targets of positive selection by PAML's branch-site models, or sites identified by DIVERGE as differing in evolutionary rate in the two families, were mapped on the 3D structure of bovine γB crystallin (PDB 1AMM; Kumaraswamy et al. 1996) using MacPyMol (Delano Scientific, San Carlos, CA).

Finally, we applied PAML's random-site, branch, and branch-site models to a data set of only γN sequences (γN-data set), with a specific focus on the fish-specific duplication event (fig. 1*b*; Wistow et al. 2005). Given the tendency of crystallins to experience adaptive conflict, a selective trade-off potentially reconcilable through gene duplication and subsequent divergence (Wistow 1993; Hughes 1994), such analyses may reveal further insight into molecular evolution in the $\beta\gamma$ lens crystallins.

Results

In order to investigate the molecular evolution of vertebrate $\beta\gamma$ lens crystallins, we assembled an alignment of 76 crystallin sequences through Genbank searches and database mining of genome projects. Five EP sequences from the Japanese firebelly newt (*C. pyrrhogaster*) were also included to serve as outgroups (Takabatake et al. 1992; Ogawa et al. 1997, 1998). Similarly to the $\beta\gamma$ lens crystallins, these EP proteins also contain two domains, each composed of two $\beta\gamma$ motifs, and are considered to be homologous but somewhat distantly related to the $\beta\gamma$ lens crystallins (Wistow et al. 1995; Shimeld et al. 2005). To this assembled data set, we

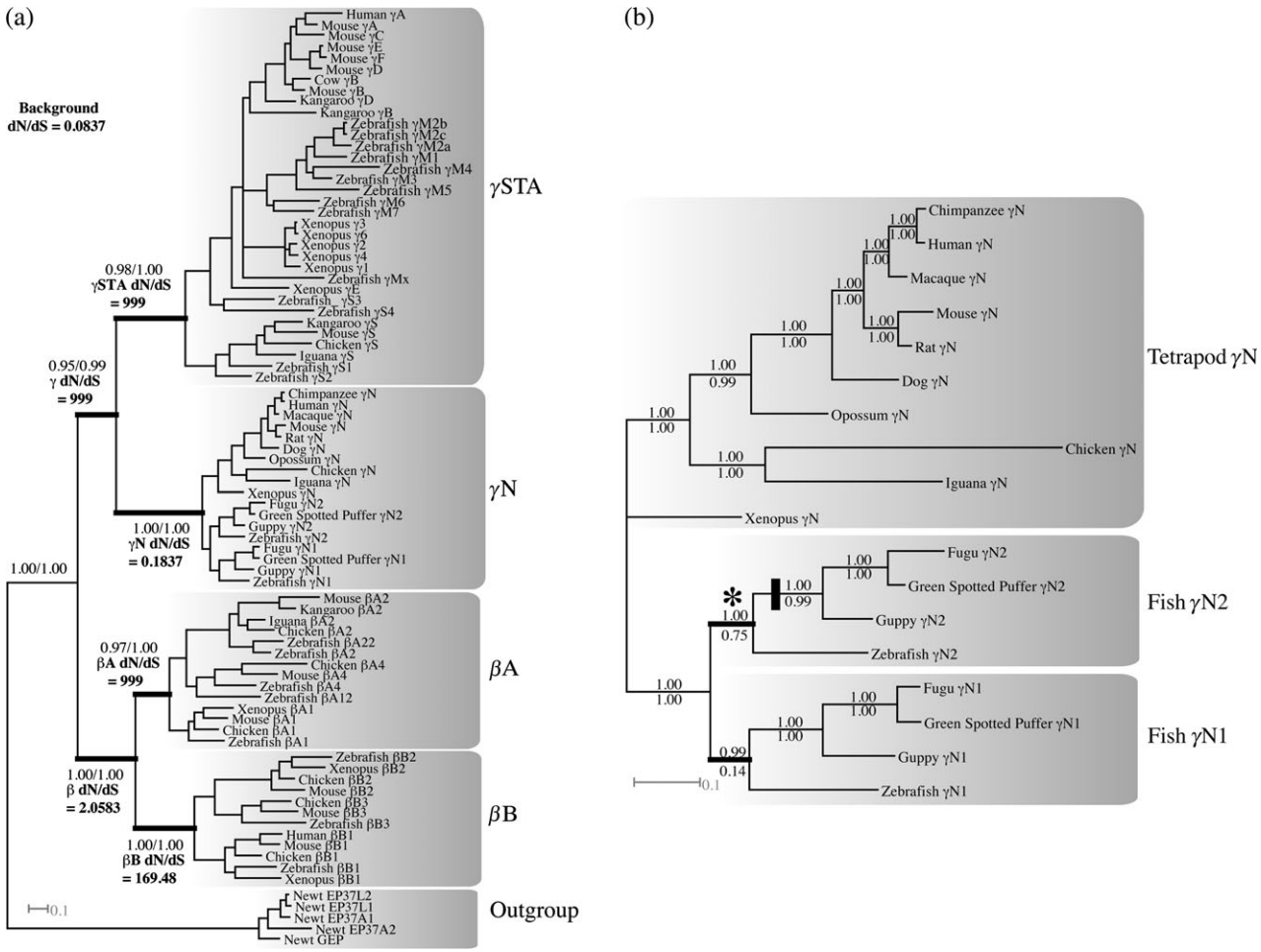

FIG. 1.—Phylogenetic estimates for the (*a*) βγ lens crystallin superfamily and (*b*) the γN crystallin family, derived from Bayesian analyses of the βγ and γN-data sets, respectively. The topologies and branch lengths are derived from codon-partitioned nucleotide analyses. Key clades are labeled, and statistical support values (PP) are shown along branches (codon-partitioned nucleotide analysis/JTT + Γ amino acid analysis), with complete support values for the βγ phylogeny provided in supplementary figure 2, Supplementary Material online. A scale bar is provided, in nucleotide substitutions per site. (*a*) Estimates of d_N/d_S for key branches (in bold) within the βγ superfamily phylogeny, derived from the 6fgb branch model (table 1), are shown alongside the relevant branches (γ, β, γN, γSTA, βA, and βB). (*b*) Thick lines indicate postduplication fish-γN branches that were tested for the signature of positive selection (see text), with positive selection being detected along the fish-γN2 branch (asterisk). The novel γN2 intron (see fig. 4) is presumed to have originated along the branch marked with a vertical bar. Phylogenetic analysis of an amino acid translated data set, under a JTT + Γ model placed the zebrafish (*Danio rerio*) γN1 crystallin sister to all other fish γN crystallins.

added complete coding sequence data for two γN sequences from the guppy, *P. reticulata*, which we obtained through the sequencing of cDNA derived from guppy head tissue (Genbank accession numbers, γN1:FJ589749, γN2:FJ589748). The putative protein-coding regions of the guppy γN sequences are both 549 nt long, as were the previously described (Wistow et al. 2005) γN sequences of the zebra fish, *D. rerio*, and contain no frameshifting mutations (fig. 2). As described above, the data sets containing only γN and γSTA crystallins (γ-alignment), or only γN crystallins (γN-alignment), represent subsets of this βγ data set (supplementary fig. 1, Supplementary Material online).

Phylogenetic Analysis

The phylogeny of the βγ lens crystallins (βγ-alignment) was estimated using a Bayesian approach. Nucleotide and amino acid analyses resulted in generally similar topologies (fig. 1*a*, supplementary fig. 2, Supplementary Material online). In both cases, the γN clade was found to nest sister to a large clade containing the γSTA crystallins. Sister to this was the βA and βB crystallins, which formed reciprocally monophyletic groups. BF comparisons generally favored this topologically unconstrained analysis over versions where the γN clade was constrained to nest either with the β crystallins (nucleotide 2*ln(BF) = 16.46 = decisive; amino acid 2*ln(BF) = 6.54 = strong), or external to all other ingroup crystallins (nucleotide 2*ln(BF) = 7.96 = strong; amino acid 2*ln(BF) = −0.38 = ambiguous). Our ability to resolve the phylogenetic location of the γN family resulted from our inclusion of several nonlens βγ-motif containing sequences of the Japanese Fire-Bellied newt (*C. pyrrhogaster*) as outgroups. Although several βγ motif–containing proteins have been described (reviewed by Piatigorsky 2007), these tend to be significantly more

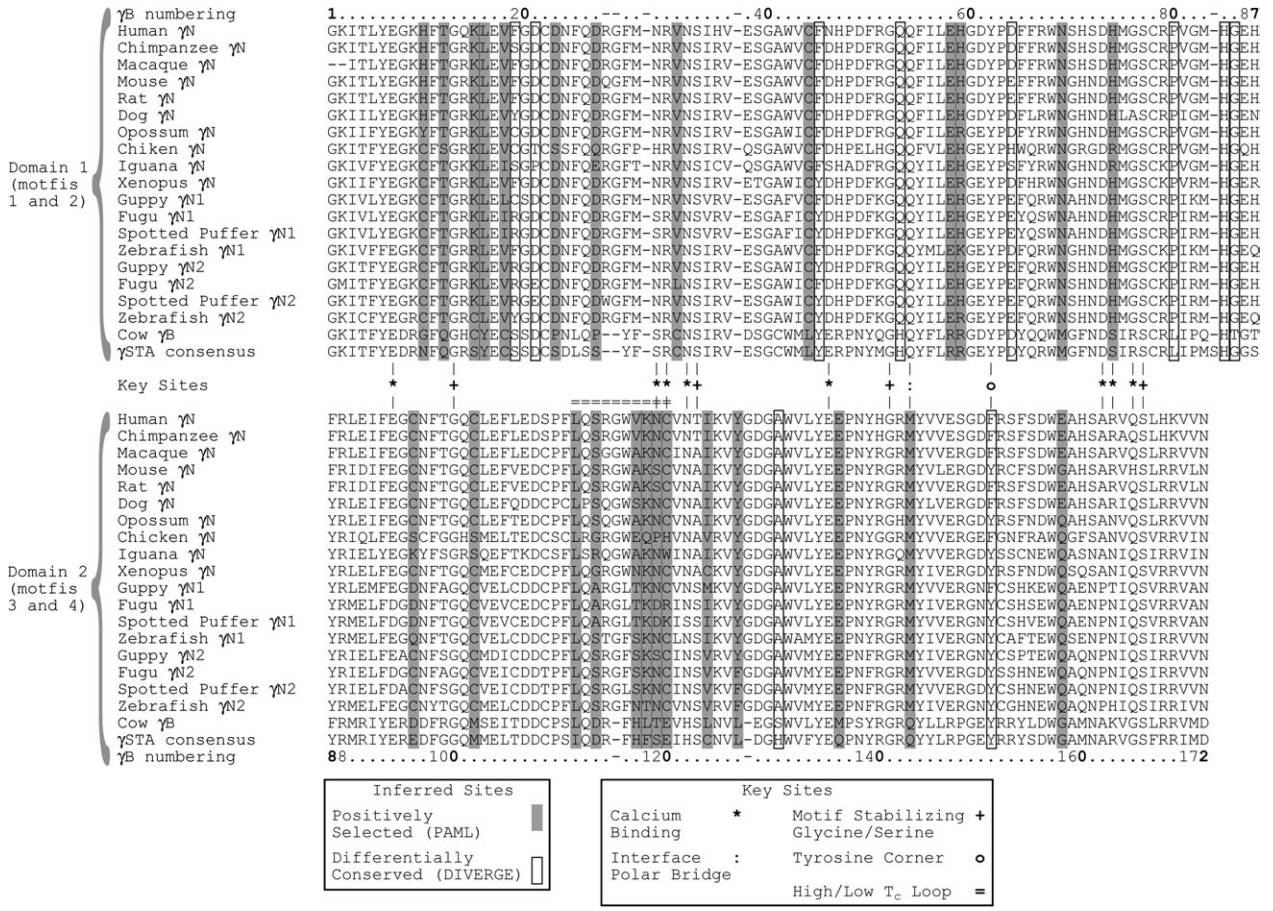

FIG. 2.—Alignment of γN crystallin amino acid sequences, with cow (*Bos taurus*) γB crystallin and a consensus sequence of γSTA crystallins. In this figure, the homologous N-terminal (top) and C-terminal (bottom) domains are aligned to each other. Sites identified by BEB analysis in PAML, under a branch-site model with the branch connecting the γN and γSTA clades, are shaded dark gray (PP > 0.95), and sites identified by DIVERGE as displaying clade-specific patterns of sequence conservation are boxed (PP > 0.90). Functionally and/or structurally important sites that are discussed in the text are indicated: "*" calcium-binding sites (Jobby and Sharma 2007); ":" interface polar bridge (Norledge et al. 1997; Purkiss et al. 2002); "+" motif stabilizing glycines and serines (Blundell et al. 1981); "o" tyrosine corner (Hemmingsen et al. 1994); "=" motif 3 loop associated with high/low T_c phase separation behavior (Norledge et al. 1997). A complete alignment, including γN, γSTA, βA, βB, and outgroup sequences, is presented in supplementary figure 1, Supplementary Material online.

divergent (Shimeld et al. 2005). The AIM1 protein (Ray et al. 1997) provides an alternative option, but although it is relatively closely related to the lens βγ crystallins, it has experienced several duplications of the βγ motif, and the ideal alignment of out and ingroup sequences is not obvious as a result.

The fact that our rooted phylogenetic analyses place the γN crystallins sister to the γSTA crystallins (fig. 1*a*) supports the hypothesis that the exon–intron structure of γN crystallins represents an intermediate state between the ancestral arrangement of β crystallins and the derived arrangement of the γ crystallins (Wistow et al. 2005). This scenario was first suggested by Wistow et al. (2005), but, because no outgroup sequences were included in this earlier study, the phylogenetic placement of the γN family among γSTA and β crystallins was equivocal. If instead our analyses placed the γN crystallins sister to the β crystallins, or sister to all other βγ lens crystallins, more complicated patterns of intron evolution would be required to explain the current distribution. Our analyses have assumed that extant βγ lens crystallins are derived from a single two-domain ancestor, a commonly accepted view (e.g., Quax-Jeuken et al. 1985; van Rens et al. 1992; Wistow et al. 2005); however, alternative hypotheses do exist. Lubsen et al. (1988) suggested that divergence between γ and β crystallins occurred at the single-domain stage, with subsequent intron loss in the γ lineage and independent duplication and fusion events occurring in both lineages. Later, D'Alessio (2002) postulated that two-domain β and γ crystallins originated independently through the fusion of single "N-terminal" and "C-terminal" domains that preexisted in forms with or without intradomain introns. However, the latter hypothesis (D'Alessio 2002) is less parsimonious than invoking a single fusion of intron-containing domains followed by two instances of intron loss (as described above, and in Wistow et al. 2005), and past phylogenetic analysis of single βγ domains (Shimeld et al. 2005) refutes the former hypothesis (Lubsen et al. 1988). The exon–intron structure in the *Cynops* EP proteins is unknown, but the arrangement found in extant β crystallins, where each βγ-motif is bordered by an intron, is presumed to be ancestral as it is also found in the relatively closely related AIM1 proteins and the distantly

related single-domain proteins of *Ciona* (Lubsen et al. 1988; de Jong et al. 1994; Ray et al. 1997; Shimeld et al. 2005).

Although the Bayesian estimates of topology showed monophyletic γN, βA, and βB clades with high PPs, the relationships were less clearly resolved among the γS, γTerrestrial, and γAquatic clades (fig. 1*a*, supplementary fig. 2, Supplementary Material online). First, both nucleotide and amino acid analyses failed to resolve a monophyletic γS clade, with the zebra fish's γS3 and γS4 crystallins branching sister to the γAquatic and γTerrestrial crystallins. This is contrary to Wistow et al.'s (2005) Neighbor-Joining phylogenetic analysis of βγ crystallins, which had similar taxonomic coverage compared with our study, though we note that the monophyly of γS sequences in their study did not have strong bootstrap support. Second, the phylogenetic analyses do not clearly resolve the relationship among the various classically defined γ crystallins (the (Terrestrial and Aquatic groups; Wistow et al. 2005). Sequences within this portion of the phylogeny tend to group by order, with mammalian-, amphibian-, and fish-derived sequences forming separate clusters, but the relationships among these clusters remain unresolved. This clustering of sequences by taxonomic group could reflect either independent, recent gene duplication events within each order, or ancestral duplications followed by homogenization by gene conversion; it has been suggested that both processes are operating in the γ crystallin clade (reviewed by Lubsen et al. 1988). These issues require further phylogenetic analysis of a more densely sampled γ crystallin data set to be conclusively resolved and will not be discussed further here, but it is for these reasons that we use the term γSTA crystallins to collectively refer to these sequences. As for the β crystallin family, our results show strong PP support for a division into βA and βB crystallins, which is consistent with past work (e.g., Berbers et al. 1984; van Rens et al. 1992). Within the βB crystallins, three well-defined clades are found that correspond to βB1, βB2, and βB3 crystallins, but relationships within the βA crystallins are not as clearly resolved, with low PP support values, especially regarding the division between βA2 and βA4.

We applied a similar Bayesian approach to estimating the γN family phylogeny as we did for the larger βγ superfamily. The tree topologies resulting from partitioned nucleotide and JTT + Γ amino acid analysis of the data set largely followed expected relationships (Cotton and Page 2002; Chen et al. 2003; Springer et al. 2004) and showed a fish-specific duplication (fig. 1*b*). However, the nucleotide and amino acid approaches differed on the topological arrangement of the multiple fish γN sequences; the nucleotide results showed two sister clades with identical species branching patterns, suggesting a gene duplication event in the ancestor of the sampled fishes, whereas the amino acid analysis placed the zebrafish γN1 as sister to all other fish γN sequences. The nucleotide topology provides a more parsimonious explanation of the fish γN gene duplication event, as it requires a single duplication, occurring prior to the diversification of sampled fishes, and no losses, whereas the amino acid tree requires two duplication events prior to the diversification of the sampled fishes, and two losses.

Molecular Evolutionary Analyses

In order to explore variation in d_N/d_S in the βγ and γN data sets, PAML (Yang 1997) was used to fit a variety of codon substitution models (Yang et al. 2000, 2005), using the tree topologies suggested by Bayesian analysis of the partitioned nucleotide data sets. For the βγ data set, the M0 codon model, which fits a single d_N/d_S parameter to the entire data set, estimated a low d_N/d_S value of 0.0826 (supplementary table 2, Supplementary Material online). Compared with M0, the M3 model, which allows for heterogeneity in d_N/d_S among sites without necessarily fitting a positively selected class, provided a much better fit to the data ($P < 0.0001$, df = 4; supplementary table 2, Supplementary Material online); with three sites classes fit to the data ($k = 3$) the M3 model revealed at least an order of magnitude variation in d_N/d_S among sites, though in all cases the d_N/d_S values were far below $d_N/d_S = 1$. None of the commonly used random-site test of positive selection (M2a–M1a, M8–M7, and M8–M8a) was significant (supplementary table 2, Supplementary Material online); in all cases, the more complex alternative model collapsed to an approximation of the null ($P = 1$, in all cases).

Following gene duplication, daughter copies are often free to accumulate previously forbidden mutations owing to reduced selective constraints; although this often leads to pseudogenization, daughter copies can also be preserved by positive selection if new mutations lead to beneficial modification of protein function (Fay and Wu 2003; Taylor and Raes 2004). With this in mind, we used PAML to apply a branch model (Yang 1998; Yang and Nielsen 1998) to the βγ-alignment in order to compare the type and strength of selection acting on the branch leading to the γN family with that found on other similarly ancient, postduplication branches. Our 6fgb model, a branch model allowing lineages leading to the ancient gene duplications to have independently estimated d_N/d_S values, indicated that d_N/d_S was much greater than one (suggesting positive selection) for all of the branches except the branch leading to the γN family (fig. 1*a*); d_N/d_S was estimated as 0.1837 for the γN branch and at least 2.0583 for the five other branches. This model fits the data significantly better than the null M0 model ($P < 0.0001$, df = 5; table 1). Though our data set spans a large evolutionary distance, with the total tree length estimated at 73.71 nt substitutions per codon, our d_N/d_S estimates seems unlikely to be affected by sequence saturation; d_S, the synonymous substitution rate, was less than 2 for all but two branches within the ingroup, $d_S = 1.06$ for the γN branch, and $d_S < 0.05$ for the five other key branches in the 6fgb model. The number of nonsynonymous changes estimated to have occurred along each of these six branches ranged from 21 to 47 (with 44 occurring along the γN branch); given the d_S estimates for these branches, these numbers suggest a rapid accumulation of amino acid substitutions on all lineages save the γN branch. We also incorporated an alternative framework for βγ lens crystallin evolution (Lubsen et al. 1988) by fitting a similar branch model (our 4fgb model) to a data set that contained no outgroup sequences, leaving the relationship between γSTA, γN, and β crystallins unresolved. Within this modified data set, the familial relationships for each domain are consistent

Table 1
Parameter Estimates and LRTs for Branch Models (PAML) Applied to the $\beta\gamma$ Data Set

Model	T_i/T_v	Background ω (d_N/d_S)	Foreground ω (d_N/d_S)	ln L	Null Model (df)	P-Value
M0	1.6023	0.0826	N/A	−26,253.0415	N/A	
6fgb	1.5781	0.0837	γN = 0.1837, γSTA = 999 γ = 999, βB = 169.4758 βA = 999, β = 2.0583	−26,234.0114	M0 (5)	$P < 0.0001$
M0 (ingroup only)	1.6385	0.06932	N/A	−24,538.6977	N/A	
4fgb (ingroup only)	1.6323	0.0707	γN = 0.0026, γSTA = 999 βB = 797.7312, βA = 999	−24,528.6099	M0 (ingroup only) (3)	$P < 0.0005$

with those predicted by Lubsen et al. (1988), and the results of our branch model analysis were qualitatively unchanged from those found using the complete data set (table 1).

Our branch model results suggest that the γN family has evolved conservatively since its origin via gene duplication, with the γSTA family having apparently experienced strong diversifying selection. Unlike γSTA crystallins, γN crystallins are known to be highly unstable and have low solubility (Wistow et al. 2005). In order to identify candidate sites that may be contributing to these, and possibly other, functional differences, we applied a branch-site model (Yang and Nielsen 2002; Zhang et al. 2005) to a reduced data set (γ-alignment), which allowed us to detect positively selected sites along the branch connecting γN and γSTA crystallins. The γ-alignment was used, rather than the larger βγ-alignment, because branch-site tests constrain all background branches to have equivalent d_N/d_S values, and the results of branch model analysis (our 6fgb model) revealed this to be a strongly violated assumption (table 1). The alternative branch-site model (Test 2 of Zhang et al. 2005) fits the data significantly better than the null model ($P = 0.007$, df = 1; table 2), suggesting strong positive selection in a class of sites along this lineage ($\omega_2 = 6.4$,). BEB assignment of sites to purifying, neutral or positively selected site classes (Yang et al. 2005) indicated that 25 sites were targets of positive selection with PP > 0.95 (table 2). The mean Grantham distances (Grantham 1974) between the inferred ancestral residues of the γN and γSTA clades was significantly greater for these 25 positively selected sites than for other sites with inferred changes (mean ± SE, 115.0 ± 12.4, and 58.1 ± 6.2, respectively; one-way t-test assuming unequal variances, $P < 0.001$; one-way Mann–Whitney U-test, $P < 0.001$). Of these 25 sites, only eight were in motifs 2 and 4, which together form the domain interface (figs. 2 and 3), and only three of the substitutions occur at sites clearly directed into the protein (either into a domain or into the domain interface) (figs. 2 and 3). Interestingly, in a few cases, the same site was identified in both domains (fig. 2).

As protein function evolves, the structural importance of particular amino acid sites may change, resulting, for example, in previously neutral and variable sites gaining importance and becoming fixed for specific amino acids. One of the ways in which functional diversification among gene families can therefore manifest itself is through altered levels of sequence conservation (Gu 1999; Gaucher et al. 2002). We searched for sites displaying the statistical signature of functional divergence (differing degrees of sequence conservation) between the γN and γSTA families using DIVERGE (Gu and Vander Velden 2002). Significant variation in amino acid substitution rate between the γN and γSTA clades was noted ($P < 0.0001$, df = 1), with 11 sites being identified as differing significantly in the degree of constraint between the two clades (PP > 0.90) (table 3). As with the positively selected sites identified by PAML, most of these substitutions were located on the surface of the protein (figs. 2 and 3). Several of the sites identified by PAML or DIVERGE (figs. 2 and 3) are known to be of functional importance in various βγ crystallins (reviewed

Table 2
Parameter Estimates and LRTs for Branch-Site Models (Test 2: PAML) Applied to the γ Data Set

Model	Parameter	Site Class (SC) SC-0	SC-1	SC-2a	SC-2b	ln L	P-Value (df)	BEB Identified Sites with PP > 0.95 (cow γB numbering)
NULL	T_i/T_v	1.7571				−15,097.2616	N/A	N/A
	SC proportion	0.6299	0.0172	0.3435	0.0094			
	Background ω (d_N/d_S)	0.0879	1	0.0879	1			
	Foreground ω (d_N/d_S)	0.0879	1	1	1			
ALT	T_i/T_v	1.7619				−15,093.6009	$P = 0.0068$ (1)	25 sites (10, 12, 15, 16, 18, 23, 27, 32, 44, 58, 59, 69, 74, 96, 102, 114, 117, 118, 119, 120, 124, 127, 126, 143, 158)
	SC proportion	0.6432	0.0176	0.3302	0.0090			
	Background ω (d_N/d_S)	0.0893	1	0.0893	1			
	Foreground ω (d_N/d_S)	0.0893	1	6.3756	6.3756			

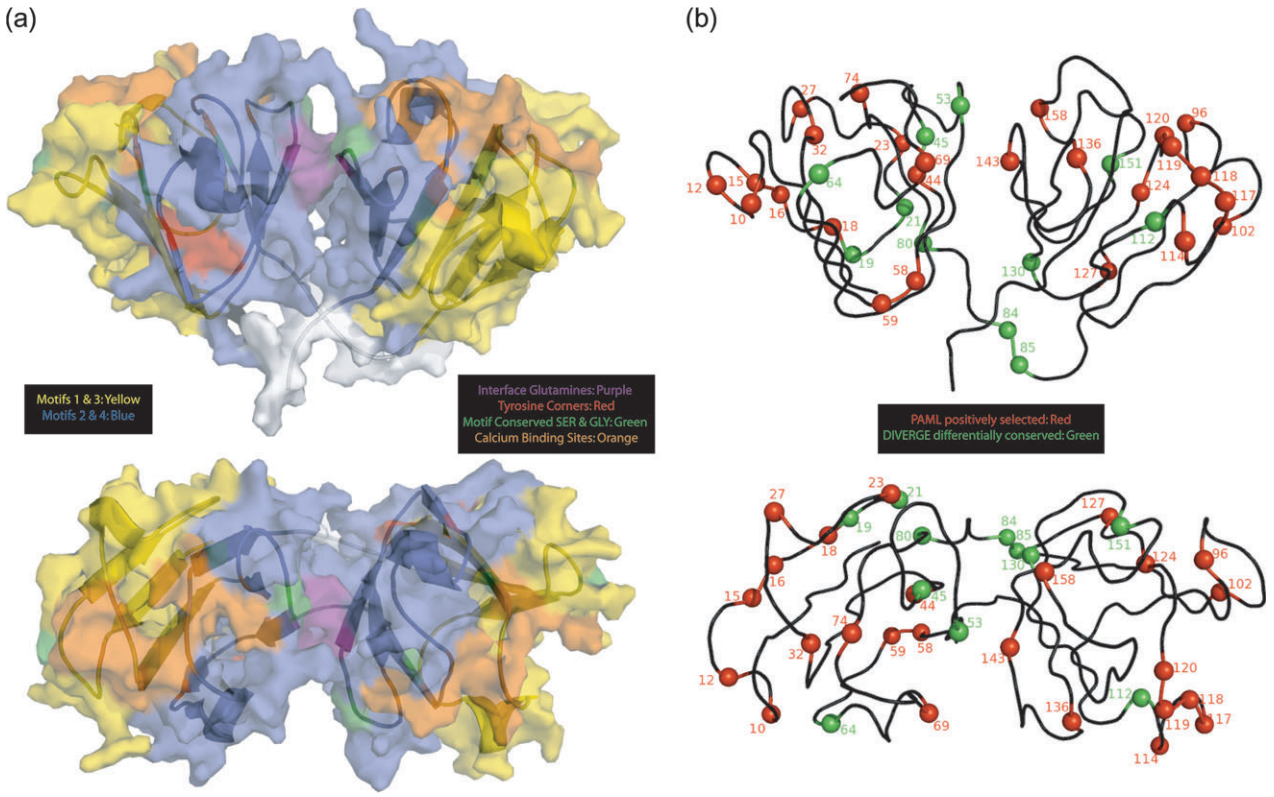

FIG. 3.—Three-dimensional structure of cow (*Bos taurus*) γB crystallin (PDB 1AMM), showing (*a*) structurally and functionally important sites and (*b*) sites inferred to be associated with functional divergence between γN and γSTA crystallins. Structural images are shown twice, once from the "side" of the protein, with the interdomain linker region located at the bottom, and again with the top of the image rotated 90° toward the viewer. (*a*) The inferred protein surface is shown superimposed over a cartoon schematic showing the various β strands and α helices that compose the four βγ motifs. Motifs 1 and 3, and Motifs 2 and 4, are shown in yellow and blue, respectively. Structurally and functionally important sites are indicated by color: interface glutamines, purple; tyrosine corners, red; motif conserved serine and glycines, green; and calcium binding sites, orange. (*b*) A ribbon diagram of cow γB with sites inferred by PAML to be under positive selection pressures along the branch connecting the γN and γSTA clades, and by DIVERGE to be differentially conserved in the two clades, indicated by red and green spheres, respectively. Sites are numbered according to position in the cow γB crystal structure.

by Bloemendal et al. 2004), and we discuss possible implications of changes at these sites in the Discussion.

In order to better explore the consequences of the γN gene duplication in fishes, a variety of codon models (Yang et al. 2000, 2005) were applied to a data set composed only of γN sequences. Similar to the βγ data set, the reduced γN data set was found to have a low overall d_N/d_S value of 0.05 (M0 model), indicative of strong purifying selection (supplementary table 3, Supplementary Material online). Although the M3–M0 test indicated variation in selection pressure among sites ($P < 0.0001$, df = 4), none of the other random-site LRTs (M2a–M1a, M8–M7, or M8–M8a) provided evidence suggesting pervasive action of positive selection (supplementary table 3, Supplementary Material online).

For the γN data set, we employed branch (Yang 1998; Yang and Nielsen 1998) and branch-site (Yang and Nielsen 2002; Zhang et al. 2005) models in order to investigate possible changes in selection pressure following the fish-specific γN duplication event. Branch models did not detect changes in selection pressure along any of the branches leading to the fish γN1 or γN2 crystallins (Fish-γN1, Fish-γN2, and γN1 and γN2 branches, $P > 0.67$ in all cases; table 4). However, allowing for clade-specific d_N/d_S estimates for both γN1 and γN2 did fit the data significantly better and suggested that γN2 sequences were evolving less conservatively than γN1 sequences (γN1 and γN2-clade; $P = 0.0445$; table 4). Similarly, branch-site analysis found statistically significant evidence for site-specific positive selection operating along the branch leading to γN2

Table 3
Parameter Estimates and LRTs from an Analysis of Functional Divergence (DIVERGE) Applied to the γ Data Set

Coefficient of Functional Divergence, θ ± SE	LRT (Null Constrains θ = 0)	*P* Value (df)	Identified Sites with PP > 0.90[a] (Bovine γB Numbering)
0.5440 ± 0.0800	46.7480	<0.0001 (1)	19, 21, 45, 53, 64, 80, 84, 85, 112, 123[a], 130, 151

[a] Site 123 was identified with PP = 0.70 but is discussed in the text.

Table 4
Parameter Estimates and LRTs for Branch Models (PAML) Applied to the γN Data Set

Model	T_i/T_v	Background ω (d_N/d_S)	Foreground ω (d_N/d_S)	ln L	Null Model (dof.)	P Value
M0	2.1604	0.0551	N/A	−4,450.11792	N/A	N/A
Fish-γN1	2.1642	0.0555	0.0362	−4,450.02947	M0 (1)	$P = 0.6741$
Fish-γN2	2.1601	0.0549	0.0702	−4,450.08487	M0 (1)	$P = 0.7911$
γN1&γN2 Branch	2.1645	0.0533	γN1-b = 0.0333, γN2-b = 0.0818	−4,449.96094	M0 (2)	$P = 0.8547$
γN1&γN2 Clade	2.1740	0.0608	γN1-c = 0.0366, γN2-c = 0.0662	−4,447.00461	M0 (2)	$P = 0.0445$

crystallins ($P = 0.0206$; table 5) but not γN1 crystallins ($P = 0.2137$; table 5). One site was identified as positively selected along the γN2 branch with a high PP (PP = 0.945 for site 70).

Novel γN Intron

Examination of the γN sequences in the pufferfish genome databases (Aparicio et al. 2002; Jaillon et al. 2004) indicated the presence of a novel intron in the first domain of the γN2 paralog. However, this novel γN2 intron was not observed in the zebrafish γN2 sequence (Wistow et al. 2005). In order to determine whether or not this novel intron is present solely in pufferfish, we used paralog-specific primers to search for this intron in the guppy, *P. reticulata*. We obtained strong single-band amplicons using both γN1 and γN2-specific primers, but the γN2 amplicon was larger than expected if only exonic sequence data were present. We then cloned and sequenced these amplicons to obtain intronic sequence data (Genbank accession number, FJ589750), which is presented in figure 4 along with the introns from pufferfish γN2. The novel intron is a short phase-2 intron (only 75 bp in the guppy and less than 200 bp in both pufferfish) and possesses the more common GT-AG splice-sites in the guppy but the more rare GC-AG splice-sites in both pufferfish. The novel intron possesses a AC microsatellite 13–14 repeats long in the two pufferfish species that is not present in the guppy. Using the guppy and pufferfish intron sequences to search both Genbank and the RepeatMasker (Smit et al. 2004) repetitive-sequence databases yielded no positive results.

Discussion

That β and γ crystallins represent different subgroups of a larger superfamily, with similar tertiary structure, was first recognized nearly three decades ago (Driessen et al. 1981; Wistow et al. 1981). Subsequent work revealed further subdivision within the β and γ crystallins, and recently an entirely new family, the γN crystallins, was discovered (Wistow et al. 2005). The γN crystallins were shown to have an intermediate genetic structure between γ and β crystallins, so it is of interest to clarify their evolutionary relationships with other βγ lens crystallins. Furthermore, they were found to have unusual biochemical properties, which raises questions as to how and why this is so. We therefore employed phylogenetic and molecular evolutionary methods to investigate patterns of sequence evolution in the βγ crystallins, with a specific focus on the newly discovered γN family.

Our phylogenetic analyses found the γN family to be sister to a clade containing all other γ crystallins, including the γS, γTerrestrial, and γAquatic crystallins (the γSTA crystallins), confirming the suggestion (Wistow et al. 2005) that its hybrid gene structure reflects an intermediate state between the ancestral β-crystallin and derived γ-crystallin arrangements. Estimates of d_N/d_S using maximum likelihood–based phylogenetic methods indicate that although there is

Table 5
Parameter Estimates and LRTs for Branch-Site Models (PAML) Applied to the γN Data Set

		Site Class (SC)					
Model	Parameter	SC-0	SC-1	SC-2a	SC-2b	ln L	P Value (df)
Fish-γN1 NULL	T_i/T_v			2.2002		−4,431.43400	N/A
	SC proportion	0.9301	0.0200	0.0489	0.0010		
	Background ω (d_N/d_S)	0.0534	1	0.0534	1		
	Foreground ω (d_N/d_S)	0.0534	1	1	1		
Fish-γN1 ALT	T_i/T_v			2.1942		−4,430.66105	$P = 0.2137$ (1)
	SC proportion	0.9553	0.0266	0.0236	0.0005		
	Background ω (d_N/d_S)	0.0533	1	0.0533	1		
	Foreground ω (d_N/d_S)	0.0533	1	10.4976	10.4976		
Fish-γN2 NULL	T_i/T_v			2.2080		−4,431.44228	N/A
	SC proportion	0.9227	0.0200	0.0562	0.0012		
	Background ω (d_N/d_S)	0.0533	1	0.0533	1		
	Foreground ω (d_N/d_S)	0.0533	1	1	1		
Fish-γN2 ALT	T_i/T_v			2.2069		−4,428.7604	$P = 0.0206$ (1)
	SC proportion	0.9436	0.0208	0.0349	0.0008		
	Background ω (d_N/d_S)	0.0540	1	0.0540	1		
	Foreground ω (d_N/d_S)	0.0540	1	357.3777	357.3777		

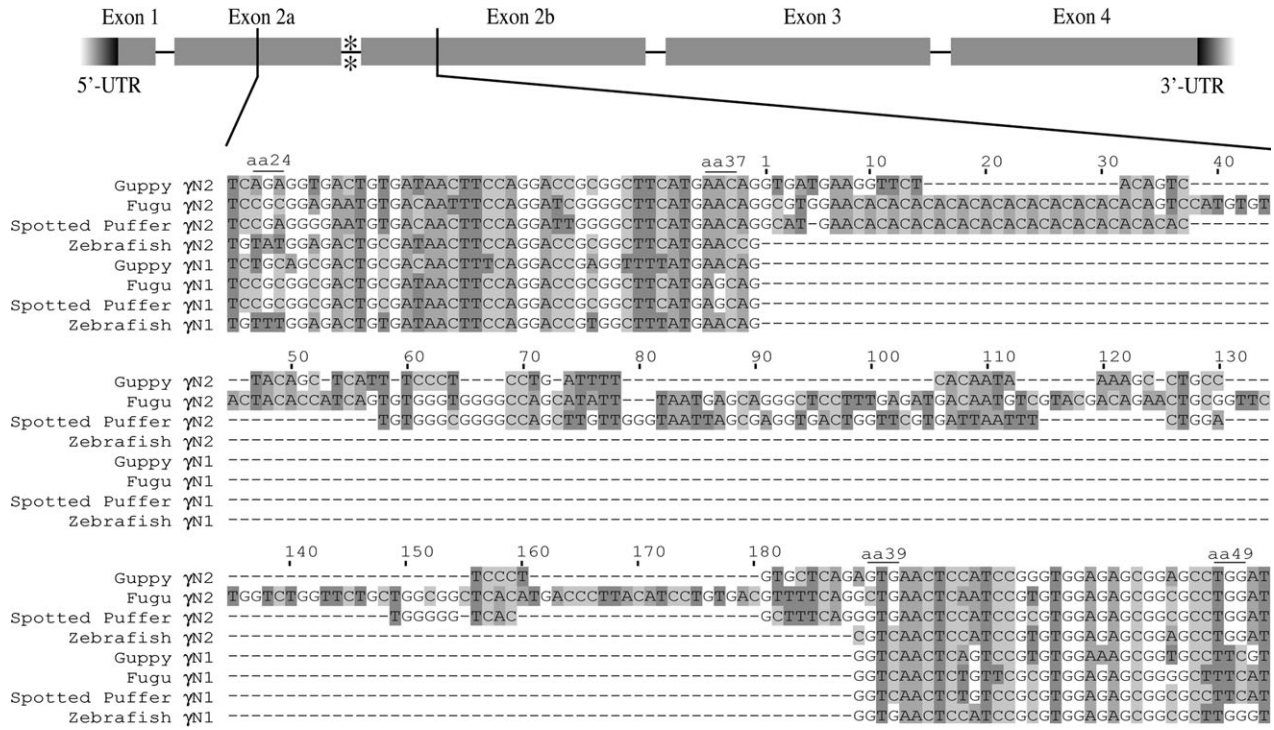

FIG. 4.—Alignment of fish γN crystallins, showing the novel γN2 crystallin intron. Top, a schematic of novel intron–containing γN2 crystallins; coding sequences (gray bands) are shown to scale, whereas untranslated regions (black bands) and introns (black lines) are not. Bottom, a nucleotide alignment of partial γN crystallins showing the distribution and sequence of the novel intron in fish γN crystallins and the adjacent coding sequence regions. Site numbering follows zebrafish γN1 and γN2 (Wistow et al. 2005). The nucleotide sequence of the guppy novel intron has been deposited in Genbank with accession number FJ589750.

evidence for some variation in d_N/d_S rates among sites, most sites are experiencing strong purifying selection throughout much of their evolutionary history, a result typical of most gene families except for those involved in reproduction or immune system coevolutionary chases (Yang and Bielawski 2000; Fay and Wu 2003). These results are consistent with previous studies indicating the predominant effects of purifying selection on crystallin gene evolution (de Jong et al. 1984; Aarts et al. 1988, 1989).

Although our results are consistent with strong purifying selection overall, lineages leading to ancient crystallin gene duplications were found to have strong positive selection, with the notable exception of the lineage leading to the γN family. This intriguing result suggests that the early history of the γN family, after it began to diverge from γSTA crystallins, was conservative and provides support for the hypothesis of a maintained ancestral function in γN crystallins (Wistow et al. 2005). Future efforts to determine the enigmatic prelens role of βγ crystallins would therefore benefit from further investigation of the γN family. Although this is not the first use of a tree-based, d_N/d_S approach to infer the maintenance, or lack thereof, of ancestral function in the context of gene family evolution (e.g., Zhang et al. 1998; Merritt and Quattro 2001), it is, to our knowledge, the first application of such methods to vertebrate lens crystallins. This is rather surprising, as crystallins have long served as key examples of the role ancestral functions have in preserving duplicate genes (Piatigorsky 2007).

The suggestion of a retained ancestral function in γN crystallins is based in part on their unusual properties (Wistow et al. 2005). In contrast to other βγ crystallins, γN crystallins have dramatically reduced stability and solubility, characteristics that are difficult to reconcile with their role in the lens, and may be indicative of an adaptive conflict with a retained ancestral function. Our analyses identified a number of sites that would serve as likely candidates for some of the striking structural and functional differences between γN and other βγ crystallins. Although a high-resolution structure of γN has yet to be determined, crystal structures of a number of other γ crystallins are known, including γA-F and the C-terminal domain of γS. The first βγ crystallin structure to be solved, cow γB, unexpectedly revealed striking similarities in structure among the four motifs, despite their high degree of sequence divergence (Blundell et al. 1981). Four highly conserved residues were proposed to be important for stabilizing the interaction between the first loop and third β-strand in each of the four motifs (Blundell et al. 1981). Using motif-1 numbering, the first (Y6), third (S34), and fourth (R36) of these sites appear to help anchor the first loop to the adjacent β-strands through hydrogen bonding, whereas the small size of the second residue (G13) permits the sharp turn made by the loop. Two of these residues were identified in our analyses, sites 45 (which corresponds to Y6 in motif 2) and 123 (which corresponds to S34 in motif 3); both are quite variable among γN crystallins.

Subsequent crystal structures of other γ crystallins also confirmed the existence of dense charge clusters, which may be important not only in mediating intermolecular contacts in the lens, but also may contribute to overall protein stability (Blundell et al. 1981; Norledge et al. 1997; Basak

et al. 2003). Sites 58, 59, 96, and 143 identified in our analyses are all part of these networks of charged residues. Mutations at many of these sites have been found to disrupt these charged clusters and cause cataracts in humans. For example, the mutation R58H is known to cause cataracts in humans, at least in part by disrupting intermolecular ion pairing with D156 and lowering the solubility of the crystallin (Basak et al. 2003). This site is variable in γ but fixed as a glutamic acid in γN crystallins. Mutations at two other sites identified in our analyses have also been linked to cataract formation in humans, due to decreased solubility (P23T, Evans et al. 2004), or increased crystallization kinetics (C18S, Asherie et al. 2001). γN crystallins are fixed for Asp at site 23 and have a variety of residues other than Cys at site 18.

A number of other residues were also identified in our analyses that are likely to affect γ crystallin structure and function. A common conformational feature of many β-barrel proteins containing Greek key motifs, including $\beta\gamma$ crystallins, tyrosine corners are thought to contribute to the stability and proper folding of these proteins (Hemmingsen et al. 1994). A residue identified in our analyses (Site 151) is typically a Tyr in $\beta\gamma$ crystallins and forms the basis of the tyrosine corner of motif 4 (Hemmingsen et al. 1994; Bloemendal et al. 2004). In γN crystallins, this site is variable, with several members instead possessing Phe, which is likely to interfere with the normal H-bonding and packing interactions for a tyrosine corner arrangement.

Among crystallins, γ-crystallins are rather unusual in that they are known to undergo a phenomenon called fluid–fluid phase separation at physiological temperatures (Broide et al. 1991). γ-Crystallins can be divided into those that are characterized by high- and low-phase separation temperatures (T_c) at which separation into two fluid phases of different protein density occurs. Thought to result from varying degrees of intermolecular interactions through short-range attractive potentials, this effect results in a reversible opacification or cloudiness of the lens. Although intriguing, it remains largely unknown why γ-crystallins display this property. Examination of the crystal structure of both high (γE, γF) and low (γB, γD) T_c proteins suggests one of the main structural differences underlying this effect is the conformation of a loop in motif 3 (residues 112–122; Norledge et al. 1997). This loop contains one of the main clusters of residues identified in our analyses (112, 114, and 117–120), with many of the residues at these sites showing differences in charge and hydrophobicity in comparisons between the γN and other $\beta\gamma$ crystallins. Structural studies of the C-terminal domain of γS crystallin also implicate site 117′s role in intermolecular interactions (Purkiss et al. 2002).

Finally, our analyses also identified several sites likely to be involved in the binding of calcium ions. Calcium binding has been documented for several members of the $\beta\gamma$ crystallin family (Sharma et al. 1989; Rajini et al. 2001; Jobby and Sharma 2007), and the exact location of calcium-binding regions have been established in several nonlens $\beta\gamma$ crystallins, such as that of the urochordate *C. intestinalis* (figs. 2 and 3; Shimeld et al. 2005). $\beta\gamma$ crystallins appear to have two calcium-binding sites per domain, each formed primarily by four residues, and three sites, one in motif 2 and two in motif 3, were identified as positively selected sites (Sites 74, 119, and 120). The first of these sites (Site 74) is variable in γSTA crystallins, with most sequences having either a serine or cysteine, whereas γN crystallins have a conserved histidine instead; this change could easily affect the binding pocket as histidine can carry a positive charge and is much bulkier than either cysteine or serine. The second of these sites (Site 119) is typically serine or arginine in γSTA, and mostly asparagine for γN sequences, whereas the third (Site 120) is typically aspartic acid or gluatmic acid in γSTA and cysteine in γN. In both cases, there are large physicochemical differences in the amino acids present in γN and γSTA crystallins, so again, it seems feasible that these changes would have functional consequences as far as calcium binding is concerned. Other positively selected sites were found to border the calcium-binding pockets (Sites 32, 118) and may have indirect effects on calcium-binding capabilities through alterations of the peptide backbone. Interestingly, some of these sites are also implicated in other aspects of $\beta\gamma$ structure and function, suggesting that the adaptive modification of one property, such as calcium binding, may affect other properties as well.

Calcium binding is one of the few nonstructural, molecular roles known for $\beta\gamma$ lens crystallins, so it is an intriguing candidate for ancestral or unknown additional functions. The specific role that calcium binding may have played in the evolutionary history of the $\beta\gamma$ crystallins is unknown, but it has been suggested that $\beta\gamma$ crystallins may play a nonrefractive role in the lens (Bhat 2004), and cataractous lenses are often found to have abnormal calcium levels (Duncan et al. 1994). Ever since homology between various lens and nonlens crystallin proteins was first found, it has been suggested that the ancestral crystallin proteins recruited for use in the lens were stress-resistant genes (de Jong et al. 1989; Wistow 1995), with α crystallins being related to heat-shock proteins (Franck et al. 2004) and $\beta\gamma$-lens crystallins seeming to share ancestry with stress-induced, microbial sporulation proteins (Wistow et al. 1985). Interestingly, calcium binding by these various microbial proteins appears to enhance protein stability (Bloemendal et al. 2004). Additional roles for $\beta\gamma$ crystallins need not relate to lens biology, however, and several members of the $\beta\gamma$ superfamily, including γN crystallins, have been found expressed in different tissues, such as nervous (Wistow et al. 2005) and muscle (Derome and Bernatchez 2006) tissues (see also Piatigorsky 2007). βB2 crystallin is expressed in rodent reproductive tissues, and a recent study found mice harboring a mutation in the βB2 crystallin gene had reduced fertility, abnormal testis development, and reduced gamete production (Duprey et al. 2007). It would be interesting to see the effects on development and physiology caused by the knocking out or induced overexpression of γN and other $\beta\gamma$ crystallins; such work will be invaluable for establishing the specific roles played by γN and $\beta\gamma$ lens crystallins, both in and out of the lens.

γN crystallins not only have atypical properties, relative to other $\beta\gamma$ crystallins, but they are also known to be expressed outside the lens (Wistow et al. 2005), suggesting a degree of multifunctionality that may be leading to adaptive conflict, a situation reminiscent of the α crystallins

(Wistow 1993; Hughes 1994). In fish, a γN duplication is known to have occurred along the lineage leading to teleost fishes, possibly as a result of the whole-genome duplication event known to have occurred along this lineage (Taylor et al. 2001). Our molecular evolutionary analyses of the fish-specific γN duplication event indicate that the two groups differ in overall selective constraint, with the γN1 clade evolving more conservatively, and reveal the action of positive selection operating along the branch leading to γN2 crystallins. These results can also be interpreted in light of a possible retained ancestral function in γN crystallins. If the hypothesis of the γN crystallins having retained ancestral function is true, the fish-specific γN gene duplication event may have enabled the fish γN2 crystallins to adapt to their functions in the lens, whereas the γN1 crystallins retained their ancestral role. This would be consistent with models of duplicate gene retention in which new duplicates escape pseudogenization if they partition an ancestral role (Wistow 1993; Hughes 1994).

Interestingly, our analyses also show that some fish γN2 crystallins acquired a novel intron. The phylogenetic distribution of this intron, which to date has only been found in pufferfish and guppy, suggests that it may have originated early in the acanthopterygian lineage, at least 190 Ma, according to recent investigations of divergence times among ray-finned fish (Steinke et al. 2006; Yamanoue et al. 2006; Hurley et al. 2007). Given that introns often play key regulatory roles (Jeffares et al. 2006), it may be that these crystallins have altered expression patterns, but whether or not adaptive processes (Carmel et al. 2007) drove the acquisition of this intron is unclear. Regardless, sequence evolution within this novel intron appears to be at least somewhat nonneutral, as both pufferfish sequences possess nonstandard splice sites at the 5′ end of the intron (GC-, instead of the more common GT- 5′ splice site) and a nearby AC microsatellite. Yeo et al. (2004) found that ACAC repeats act as splicing enhancers in fish introns, especially when found adjacent to nonstandard 5′ splice sites. Which appeared first in this novel intron, the GC- 5′ splice site, or the AC microsatellite, is unknown, but the fact that neither is found in the guppy reinforces the postulated link between the two intron elements.

The results of our analyses of the γN as well as the larger βγ superfamily of crystallins suggest that the γN family retained an ancestral function following its split from the γSTA group, but that it may have undergone more recent periods of gene duplication and adaptive diversification in fish. Our study has identified a number of candidate sites for explaining functional diversification within the βγ lens crystallin superfamily. The effects of mutating these candidate sites on crystallin properties such as stability, solubility, and calcium binding, will benefit from further investigation through techniques such as site-directed mutagenesis. Similarly, the relevance of the AC microsatellite in the novel introns of pufferfish, and the presumed effect of variation in microsatellite length on intron splicing, could be explored using *in vivo* splicing reporter assays (e.g., Miles et al. 2003). In terms of molecular evolutionary analyses, future work on structure–function relationships and regulatory patterns of βγ crystallins would benefit from recent methodological developments in searching for patterns of residue coevolution and selection on synonymous sites (e.g., Fares and Travers 2006; Yang and Nielsen 2008). Given how widespread protein multifunctionality appears to be (Piatigorsky 2007) and its potential importance in duplicate gene retention (Wistow 1993; Hughes 1994; Lynch and Force 2000; Bergthorsson et al. 2007), one interesting avenue of research could involve study of how, evolutionarily (e.g., Johnston et al. 2007) and mechanistically (e.g., Copley 2003), protein multifunctionality first arose. Likewise, expanded taxonomic sampling will aid in fully revealing the evolutionary history of the βγ crystallins and its many families. Possible avenues include more accurately determining the timing of the γN duplication in fish and the origin of the γN2 novel intron, or carrying out broad surveys of crystallin diversity in basal vertebrates, such as the agnathans. Such work, especially when combined with functional assays of protein function, will undoubtedly benefit efforts to determine the nonlens role of βγ crystallins and provide a useful system for studying the structural and functional consequences of gene duplication.

Supplementary Material

Supplementary tables 1–3 and supplementary figure 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to acknowledge Helen Rodd for her support and advice throughout all stages of this project. We would also like to thank Francesco Santini and Steve Walker for helpful advice on Bayesian statistics and Yogendra Sharma for answering questions on crystallin calcium binding. Helen Rodd, Asher Cutter, the members of the Chang Lab, and three anonymous reviewers provided helpful comments on this manuscript. This research was supported by a Natural Sciences and Engineering Council of Canada (NSERC) Discovery grant (BSWC) an Early Researcher Award (BSWC), an NSERC Post Graduate Scholarship (C.J.W.), and a University Health Network (University of Toronto) Vision Science Research Program fellowship (C.J.W.).

Literature Cited

Aarts HJ, den Dunnen JT, Leunissen J, Lubsen NH, Schoenmakers JG. 1988. The gamma-crystallin gene families: sequence and evolutionary patterns. J Mol Evol. 27:163–172.

Aarts HJ, Jacobs EH, van Willigen G, Lubsen NH, Schoenmakers JG. 1989. Different evolution rates within the lens-specific beta-crystallin gene family. J Mol Evol. 28: 313–321.

Aparicio S, Chapman J, Stupka E, et al. (40 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science. 297:1301–1310.

Asherie N, Pande J, Pande A, Zarutskie JA, Lomakin J, Lomakin A, Ogun O, Stern LJ, King J, Benedek GB. 2001. Enhanced crystallization of the Cys18 to Ser mutant of bovine gammaB crystallin. J Mol Biol. 314:663–669.

Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J. 2003. High-resolution X-ray crystal structures of human gammaD crystallin (1.25 A) and the R58H mutant (1.15 A) associated with aculeiform cataract. J Mol Biol. 328:1137–1147.

Berbers GA, Hoekman WA, Bloemendal H, de Jong WW, Kleinschmidt T, Braunitzer G. 1984. Homology between the primary structures of the major bovine beta-crystallin chains. Eur J Biochem. 139:467–479.

Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. Proc Natl Acad Sci USA. 104:17004–17009.

Bhat SP. 2004. Transparency and non-refractive functions of crystallins–a proposal. Exp Eye Res. 79:809–816.

Bloemendal H, de Jong W, Jaenicke R, Lubsen NH, Slingsby C, Tardieu A. 2004. Ageing and vision: structure, stability and function of lens crystallins. Prog Biophy Mol Biol. 86: 407–485.

Blundell T, Lindley P, Miller L, Moss D, Slingsby C, Tickle I, Turnell B, Wistow G. 1981. The molecular structure and stability of the eye lens: x-ray analysis of gamma-crystallin II. Nature. 289:771–777.

Broide ML, Berland CR, Pande J, Ogun OO, Benedek GB. 1991. Binary-liquid phase separation of lens protein solutions. Proc Natl Acad Sci USA. 88:5660–5664.

Brown JM, Lemmon AR. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst Biol. 56:643–655.

Carmel L, Rogozin IB, Wolf YI, Koonin EV. 2007. Evolutionarily conserved genes preferentially accumulate introns. Genome Res. 17:1045–1050.

Chen WJ, Bonillo C, Lecointre G. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol Phylogenet Evol. 26:262–288.

Copley SD. 2003. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. Curr Opin Chem Biol. 7:265–272.

Cotton JA, Page RD. 2002. Going nuclear: gene family evolution and vertebrate phylogeny reconciled. Proc Biol Sci. 269: 1555–1561.

D'Alessio G. 2002. The evolution of monomeric and oligomeric betagamma-type crystallins. Facts and hypotheses. Eur J Biochem. 269:3122–3130.

Danforth BN, Sipes S, Fang J, Brady SG. 2006. The history of early bee diversification based on five genes plus morphology. Proc Natl Acad Sci USA. 103:15118–15123.

Darwin C. 1859. On the origin of species. Cambridge (MA): Harvard University Press.

de Jong WW, Hendriks W, Mulders JW, Bloemendal H. 1989. Evolution of eye lens crystallins: the stress connection. Trends Biochem Sci. 14:365–368.

de Jong WW, Lubsen NH, Kraft HJ. 1994. Molecular evolution of the eye lens. Prog Ret Eye Res. 13:391–442.

de Jong WW, Zweers A, Versteeg M, Nuy-Terwindt EC. 1984. Primary structures of the alpha-crystallin A chains of twenty-eight mammalian species, chicken and frog. Eur J Biochem. 141:131–140.

Derome N, Bernatchez L. 2006. The transcriptomics of ecological convergence between 2 limnetic coregonine fishes (Salmonidae). Mol Biol Evol. 23:2370–2378.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature. 454:762–765.

Dohrmann M, Voigt O, Erpenbeck D, Worheide G. 2006. Non-monophyly of most supraspecific taxa of calcareous sponges (Porifera, Calcarea) revealed by increased taxon sampling and partitioned Bayesian analysis of ribosomal DNA. Mol Phylogenet Evol. 40:830–843.

Driessen HP, Herbrink P, Bloemendal H, de Jong WW. 1981. Primary structure of the bovine beta-crystallin Bp chain. Internal duplication and homology with gamma-crystallin. Eur J Biochem. 121:83–91.

Duncan G, Williams MR, Riach RA. 1994. Calcium, cell signaling and cataract. Prog Retinal Eye Res. 13:623–652.

Duprey KM, Robinson KM, Wang Y, Taube JR, Duncan MK. 2007. Subfertility in mice harboring a mutation in betaB2-crystallin. Mol Vis. 13:366–373.

Evans P, Wyatt K, Wistow GJ, Bateman OA, Wallace BA, Slingsby C. 2004. The P23T cataract mutation causes loss of solubility of folded gammaD-crystallin. J Mol Biol. 343:435–444.

Fares MA, Travers SA. 2006. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. Genetics. 173:9–23.

Fay JC, Wu CI. 2003. Sequence divergence, functional constraint, and selection in protein evolution. Annu Rev Genomics Hum Genet. 4:213–235.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17:368–376.

Fernald RD. 2006. Casting a genetic light on the evolution of eyes. Science. 313:1914–1918.

Franck E, Madsen O, van Rheede T, Ricard G, Huynen MA, de Jong WW. 2004. Evolutionary diversity of vertebrate small heat shock proteins. J Mol Evol. 59:792–805.

Frohman MA, Dush MK, Martin GR. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA. 85:8998–9002.

Gaucher EA, Gu X, Miyamoto MM, Benner SA. 2002. Predicting functional divergence in protein evolution by site-specific rate shifts. Trends Biochem Sci. 27:315–321.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. Science. 185:862–864.

Gu X. 1999. Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol. 16:1664–1674.

Gu X, Vander Velden K. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics. 18:500–501.

Hall BG. 2008. How well does the HoT score reflect sequence alignment accuracy? Mol Biol Evol. 25:1576–1580.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22:160–174.

Hemmingsen JM, Gernert KM, Richardson JS, Richardson DC. 1994. The tyrosine corner: a feature of most Greek key beta-barrel proteins. Protein Sci. 3:1927–1937.

Holder M, Lewis PO. 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat Rev Genet. 4:275–284.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci. 256:119–124.

Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang ZH, Thomas MG, Coates MI. 2007. A new time-scale for ray-finned fish evolution. Proc R Soc B-Biol Sci. 274:489–498.

Jaillon O, Aury JM, Brunet F, et al. (60 co-authors). 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature. 431: 946–957.

Jeffares DC, Mourier T, Penny D. 2006. The biology of intron gain and loss. Trends Genet. 22:16–22.

Jobby MK, Sharma Y. 2007. Calcium-binding to lens betaB2- and betaA3-crystallins suggests that all beta-crystallins are calcium-binding proteins. FEBS J. 274:4135–4147.

Johnston CR, O'Dushlaine C, Fitzpatrick DA, Edwards RJ, Shields DC. 2007. Evaluation of whether accelerated protein evolution in chordates has occurred before, after, or simultaneously with gene duplication. Mol Biol Evol. 24: 315–323.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.
Kass RE, Raftery AE. 1995. Bayes factors. J Am Stat Assoc. 90:773–795.
Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature. 267:275–276.
Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.
Kumaraswamy VS, Lindley PF, Slingsby C, Glover ID. 1996. An eye lens protein-water structure: 1.2 A resolution structure of gammaB-crystallin at 150 K. Acta Crystallogr D Biol Crystallogr. 52:611–622.
Land MF, Fernald RD. 1992. The evolution of eyes. Annu Rev Neurosci. 15:1–29.
Landan G, Graur D. 2007. Heads or tails: a simple reliability check for multiple sequence alignments. Mol Biol Evol. 24:1380–1383.
Lubsen NH, Aarts HJ, Schoenmakers JG. 1988. The evolution of lenticular proteins: the beta- and gamma-crystallin super gene family. Prog Biophys Mol Biol. 51:47–76.
Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics. 154:459–473.
Merritt TJ, Quattro JM. 2001. Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. Genetics. 159:689–697.
Miles CG, Rankin L, Smith SI, Niksic M, Elgar G, Hastie ND. 2003. Faithful expression of a tagged Fugu WT1 protein from a genomic transgene in zebrafish: efficient splicing of pufferfish genes in zebrafish but not mice. Nucleic Acids Res. 31:2795–2802.
Norledge BV, Hay RE, Bateman OA, Slingsby C, Driessen HP. 1997. Towards a molecular understanding of phase separation in the lens: a comparison of the X-ray structures of two high T_c gamma-crystallins, gammaE and gammaF, with two low T_c gamma-crystallins, gammaB and gammaD. Exp Eye Res. 65:609–630.
Nylander JAA. 2004. MrModeltest v2. Evolutionary Biology Centre, Uppsala University.Program distributed by the author
Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL. 2004. Bayesian phylogenetic analysis of combined data. Syst Biol. 53:47–67.
Ogawa M, Takabatake T, Takahashi TC, Takeshima K. 1997. Metamorphic change in EP37 expression: members of the betagamma-crystallin superfamily in newt. Dev Genes Evol. 206:417–424.
Ogawa M, Takahashi TC, Takabatake T, Takeshima K. 1998. Isolation and characterization of a gene expressed mainly in the gastric epithelium, a novel member of the ep37 family that belongs to the betagamma-crystallin superfamily. Dev Growth Differ. 40:465–473.
Piatigorsky J. 2007. Gene sharing and evolution: the diversity of protein functions. Cambridge (MA): Harvard University Press.
Posada D, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. Syst Biol. 50:580–601.
Purkiss AG, Bateman OA, Goodfellow JM, Lubsen NH, Slingsby C. 2002. The X-ray crystal structure of human gamma S-crystallin C-terminal domain. J Biol Chem. 277: 4199–4205.
Quax-Jeuken Y, Driessen H, Leunissen J, Quax W, de Jong W, Bloemendal H. 1985. Beta s-Crystallin: structure and evolution of a distinct member of the beta gamma-superfamily. Embo J. 4:2597–2602.
Rajini B, Shridas P, Sundari CS, Muralidhar D, Chandani S, Thomas F, Sharma Y. 2001. Calcium binding properties of gamma-crystallin: calcium ion binds at the Greek key beta gamma-crystallin fold. J Biol Chem. 276:38464–38471.
Ray ME, Wistow G, Su YA, Meltzer PS, Trent JM. 1997. AIM1, a novel non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. Proc Natl Acad Sci USA. 94:3229–3234.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Sassi SO, Braun EL, Benner SA. 2007. The evolution of seminal ribonuclease: pseudogene reactivation or multiple gene inactivation events? Mol Biol Evol. 24:1012–1024.
Sawyer SL, Wu LI, Emerman M, Malik HS. 2005. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci USA. 102:2832–2837.
Sharma Y, Rao CM, Narasu ML, Rao SC, Somasundaram T, Gopalakrishna A, Balasubramanian D. 1989. Calcium ion binding to delta- and to beta-crystallins. The presence of the "EF-hand" motif in delta-crystallin that aids in calcium ion binding. J Biol Chem. 264:12794–12799.
Shimeld SM, Purkiss AG, Dirks RP, Bateman OA, Slingsby C, Lubsen NH. 2005. Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. Curr Biol. 15:1684–1689.
Smit AFA, Hubley R, Green P. 2004. RepeatMasker Open-3.0. http://www.repeatmasker.org.
Springer MS, Stanhope MJ, Madsen O, de Jong WW. 2004. Molecules consolidate the placental mammal tree. Trends Ecol Evol. 19:430–438.
Steinke D, Salzburger W, Meyer A. 2006. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. J Mol Evol. 62:772–784.
Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol. 20:18–20.
Sweeney AM, Des Marais DL, Ban YE, Johnsen S. 2007. Evolution of graded refractive index in squid lenses. J R Soc Interf. 4:685–698.
Takabatake T, Takahashi TC, Takeshima K. 1992. Cloning of an epidermis-specific Cynops cDNA from Neurula library. Dev Growth Differ. 34:277–283.
Tavare S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect Math Life Sci. 17: 57–86.
Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet. 38:615–643.
Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci. 356:1661–1679.
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.
van Rens GL, de Jong WW, Bloemendal H. 1992. A superfamily in the mammalian eye lens: the beta/gamma-crystallins. Mol Biol Rep. 16:1–10.

Weadick CJ, Chang BS. 2007. Long-wavelength sensitive visual pigments of the guppy (*Poecilia reticulata*): six opsins expressed in a single individual. BMC Evol Biol. 7(Suppl. 1): S11.
Wistow G. 1993. Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem Sci. 18:301–306.
Wistow G. 1995. Molecular biology and evolution of crystallins: gene recruitment and multifunctional proteins in the eye lens. Austin (TX): R.G. Landes.
Wistow G, Jaworski C, Rao PV. 1995. A non-lens member of the beta gamma-crystallin superfamily in a vertebrate, the amphibian Cynops. Exp Eye Res. 61:637–639.
Wistow G, Slingsby C, Blundell T, Driessen H, De Jong W, Bloemendal H. 1981. Eye-lens proteins: the three-dimensional structure of beta-crystallin predicted from monomeric gamma-crystallin. FEBS Lett. 133:9–16.
Wistow G, Summers L, Blundell T. 1985. *Myxococcus xanthus* spore coat protein S may have a similar structure to vertebrate lens beta gamma-crystallins. Nature. 315:771–773.
Wistow G, Wyatt K, David L, Gao C, Bateman O, Bernstein S, Tomarev S, Segovia L, Slingsby C, Vihtelic T. 2005. gamma N-crystallin and the evolution of the beta gamma-crystallin superfamily in vertebrates. FEBS J. 272:2276–2291.
Yamanoue Y, Miya M, Inoue JG, Matsuura K, Nishida M. 2006. The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: tetraodontiformes) and divergence time estimation among model organisms in fishes. Genes Genet Syst. 81:29–39.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 15:568–573.
Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 39:306–314.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.
Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–503.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 19:908–917.
Yang Z, Nielsen R. 2008. Mutation–selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol Biol Evol. 25:568–579.
Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol. 46: 409–418.
Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 155:431–449.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.
Yeo G, Hoon S, Venkatesh B, Burge CB. 2004. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. Proc Natl Acad Sci USA. 101:15700–15705.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22:2472–2479.
Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA. 95:3708–3713.
Zharkikh A. 1994. Estimation of evolutionary distances between nucleotide sequences. J Mol Evol. 39:315–329.

Adriana Briscoe, Associate Editor

Accepted February 11, 2009