



Published in final edited form as:
Mol Vis. ; 12: 1565–1585.

Interphotoreceptor retinoid-binding protein gene structure in tetrapods and teleost fish

John M. Nickerson¹, Ruth A. Frey², Vincent T. Ciavatta¹, and Deborah L. Stenkamp²

¹ Ophthalmology Department, Emory University, Atlanta, GA

² Department of Biological Sciences, University of Idaho, Moscow, ID

Abstract

Purpose—The interphotoreceptor retinoid-binding protein (IRBP) gene possesses an unusual structure, encoding multiple Repeats, each consisting of about 300 amino acids. Our goals were to gain insight into the function of IRBP, and to test the current model for the evolution of IRBP, in which Repeats were replicated from a simpler ancestral gene.

Methods—We employed a bioinformatics approach to analyze IRBP loci in recently completed or near-complete genome sequences of several vertebrates and nonvertebrate chordates. IRBP gene expression in zebrafish was evaluated by reverse transcriptase PCR (RT-PCR) and in situ mRNA hybridizations with gene-specific probes.

Results—Patterns of exons and introns in the IRBP genes of tetrapods were highly similar, as were predicted amino acid sequences and Repeat structures. IRBP gene structure in teleost fish was more variable, and we report a new gene structure for two species, the Japanese puffer fish (*Takifugu rubripes*) and the zebrafish (*Danio rerio*). These teleost genomes contain a two-gene IRBP locus arranged head-to-tail in which the first gene, Gene 1, is intronless and contains a single large exon encoding three complete Repeats. It is followed by a second gene, Gene 2, which corresponds to the previously reported gene consisting of two Repeats spread across four exons and three introns. Each of the two zebrafish genes is transcribed. Gene 2 is expressed in the photoreceptors and RPE, and Gene 1 is expressed in the inner nuclear layer and weakly in the ganglion cell layer.

Conclusions—The tetrapod IRBP gene structure is highly conserved while the teleost fish gene structure was a surprise: It appears to be a two-gene locus with distinct Repeat organization in each open reading frame. This gene structure and gene expression data are consistent with possible neofunctionalization or sub-function partitioning of Gene 1 and Gene 2 in the zebrafish. We suggest that the two-gene locus in teleost fish arose as a consequence of either the known whole genome duplication or single gene tandem duplication.

Vertebrate interphotoreceptor retinoid-binding protein (IRBP) is the most abundant soluble protein found in the interphotoreceptor space (IPS) between photoreceptors and the retinal pigmented epithelium (RPE) [1]. It has been proposed that IRBP functions in this space to solve unique difficulties in maintaining retinoid isomerization and chemical form while retinoids cross back and forth between the RPE cell and the photoreceptor cell (PhR) [2,3]. This hypothesis is consistent with the corresponding lack of an IPS, the use of a different retinoid isomerization approach, and the absence of IRBP in invertebrates.

IRBP is expressed in developing PhRs, being turned on embryonically in the mouse at a low level [4]. Then, at birth, IRBP expression is markedly increased, and the mRNA for IRBP

rapidly accumulates, preceding the activation of many photoreceptor-specific genes [5]. A similar pattern is found in bovine eyes [6]. In zebrafish, IRBP mRNA is detected in the retina at 50 h post fertilization (hpf) [7]. In zebrafish, the pattern of cell type expression is different from that of other vertebrates, with IRBP mRNA and protein detected in both PhRs and RPE cells [7]. In PhRs, IRBP mRNA precedes the expression of cone opsins and rod opsin [7]. This precocious expression of IRBP, before many vision-specific proteins, is consistent with a possible role for IRBP in developmental, cell survival, or maintenance in the visual system [8], although it may simply reflect differential gene transcription requirements [9,10].

It is not clear how IRBP functions in the visual system, although its retinoid [11–13] and fatty acid binding [14–16] properties are well characterized in experiments using human, bovine, and *Xenopus* material [14,15,17–23]. The absence of IRBP has no apparent deleterious effect (but see Pugh and Lamb [24]) on the rate of retinoid shuttling between the RPE and the photoreceptor cell, an essential process known as the Visual Cycle [8,25]. This is not to imply that IRBP knockout mice are normal. On the contrary, the absence of IRBP yields a visual system with reduced a-wave magnitude by electroretinography, corresponding proportionally to a histologically reduced thickness of the ONL. Initially, IRBP knockout mice appear to have half the a-wave signal and about half the number of rod photoreceptor cells [8,25]. The outer segments (OS) appear disorganized, and small vacuoles are found between OS [25]. These knockout mouse studies suggest a function for IRBP in the development and maintenance of the PhR.

The mammalian IRBP gene has an unusual gene structure [26,27]. This structure has provoked interest in its evolution from a predicted simpler ancestral gene implied by the existence of distantly related protein family members. The sequence of the IRBP gene has been used to study the phylogeny of the vertebrates [28,29], and as a consequence, part of the gene sequence of IRBP is known in many (more than 600) species. In mammals, the IRBP protein contains four Repeats, each consisting of about 300 amino acids, and the three-dimensional (3D) structure of this unit, the 300 amino acid long Repeat has been solved [30,31]. A comparison of this structure to other 3D structures revealed that the IRBP Repeat is a member of a large family of proteins including enoyl-coenzyme A (CoA) hydratase [32], dienoyl CoA isomerase [33], 4-chlorobenzoyl CoA dehalogenase [34], and C-terminal protease [35]. The comparison of the primary sequences of IRBP Repeats and other family members shows weak but statistically significant similarities [36]. These family members are synthesized as monomeric polypeptides having only a single Repeat. These monomeric polypeptides can form quaternary structures of three or six polypeptides to provide a functional protein. Many of the family members act as enzymes that modify or digest hydrophobic molecules. It is not understood why there are four Repeats in a single mammalian IRBP polypeptide chain. Any enzymatic activity of IRBP remains unknown, but it probably does not include general protease activity [37] or activities with acyl-CoA substrates [38].

To begin to understand how IRBP functions, we sought to examine variably in the gene structure, with the rationale that if the gene structure varied widely, then the fundamental functional unit for vision within IRBP might correspond to a single part or subset of the full protein. Simpler IRBP orthologs might also highlight which of the protein components is the functional unit in the visual system.

An important concept in evolution [39] is the principle that following gene duplication, both copies of duplicate genes tend to be retained if multiple functions of the protein are subdivided between the duplicates. This partitioning of functions between the two genes is called sub-functionalization [40], and it is often detected in the teleost (bony) fish, which underwent a whole genome duplication (WGD) prior to or coincident with the great radiation of the teleost fish about 350 Mya [41,42]. Under this model, we hypothesize that two IRBP-like genes, each

with a different gene structure, different promoter elements, and perhaps a different set of Repeat(s), would exist in teleost genomes. We would further expect a different spatiotemporal expression profile for IRBP gene duplicates. In this study, we sought to test the potential of sub-functionalization in the IRBP locus of teleosts by examining gene structure and putative differential gene expression.

Our approach was to employ bioinformatics to analyze recently completed or near-complete genomes of several tetrapods, teleost fish, and two nonvertebrate chordates. We compared gene structure, predicted protein structure, and used these comparisons to evaluate the current model for IRBP gene evolution [26].

Here we report a high degree of conservation of IRBP gene and protein structure among the tetrapods, but also report a new locus structure for the IRBP gene in two teleost fish: Japanese puffer fish (*Takifugu rubripes*) and the zebrafish (*Danio rerio*). These two species share a pattern of exons and introns differing from that predicted for the IRBP gene in the teleost fish. Furthermore, each species contains a two-gene locus, opening the possibility of sub-function partitioning or neofunctionalization [39]. Consistent with this hypothesis, we demonstrate differences in the temporal, spatial, and cell-type expression of the two different IRBP genes in zebrafish.

METHODS

Identification and analyses of IRBP loci

Recently, the complete or nearly complete genome sequences of human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), domestic dog (*Canis familiaris*), domestic cow (*Bos taurus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), the western clawed frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), pufferfish (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), and of two urochordates (*Ciona intestinalis* and *Ciona savignyi*), have become publicly available. In addition, partial genome information is available for Rhesus monkey (*Macaca mulatta*), African elephant (*Loxodonta africana*), domestic cat (*Felis catus*), domestic sheep (*Ovis aries*), and domestic pig (*Sus scrofa*). Expressed sequence tag (EST) profiles are available for goldfish (*Carasius auratus*), threespine stickleback (*Gasterosteus aculeatus*), and fathead minnow (*Pimephales promelas*). The public databases used in this study include GenBank (National Center for Biotechnology Information, Bethesda, MD), UCSC Genome Browser (UCSC Genome Bioinformatics Group, University of California, Santa Cruz), Wellcome Trust Sanger Institute (Wellcome Trust Genome Campus, Hinxton, Cambridge, UK), the National Institute of Genetics, Mishima, Japan, and The Broad Institute, Cambridge, MA. GenBank accession or scaffold numbers for the sequences used in this article are provided in Table 1.

We analyzed IRBP gene orthologs in each of the above species with a combination of commercial software including MacVector (TM) versions 7.0 and 9.0 (Accelrys Software, San Diego, CA), and the Vector NTI suite (build 194, Invitrogen Corporation, Carlsbad, CA) implemented on a Macintosh (Apple Computer, Cupertino, CA) desktop computer, and several web-based packages (each of which is specified in the figure legends or the body of the text).

Cross-species identification of IRBP gene orthologs was based on TBLASTN searches [43] using the human IRBP amino acid sequence as the query. In most cases, the BLOSUM62 matrix and a cutoff E-value of 0.01 were employed. The gene sequence locus was bounded on 5' and 3' ends by identification of genes immediately upstream and downstream of IRBP (usually GDF2 and Annexin 8, respectively). If there were more short exons upstream or downstream of the bounding genes, these would not be considered in the analysis of the individual locus. The likelihood of large (or even small) IRBP gene fragments beyond the

immediate vicinity of the present analyses seems unlikely as no other strong sequence similarities were found by Blast searches of nonredundant or species-specific genomic DNA or EST databases. This is supported by previous Southern blot analyses suggesting that the IRBP gene is found in a single locus in all examined species [44].

The Pustell protein and DNA dot matrix programs from MacVector were used to map the approximate positions of intron-exon boundaries of newly identified IRBP genes. Closer inspection of the DNA sequences applying the consensus splice sites [45] were used to identify the precise boundaries. In most cases the splice site boundaries were confirmed by comparison to cDNA sequences from companion public databases.

The AUGUSTUS program [46] was used to predict all protein encoding genes within IRBP loci. Pairwise amino acid sequence comparisons were performed using BLASTP of the Biology Workbench, version 3.2, implemented at the San Diego Supercomputer Center, University of California, San Diego, CA Workbench. Matches to subsequence motifs, such as glycosylation sites, were identified by ad hoc grep pattern searching in a text processor (BBEdit, version 7.1.4; Bare Bones Software, Inc., Bedford, MA).

Reverse transcriptase-coupled polymerase chain reaction

Reverse transcriptase-coupled Polymerase Chain Reaction (RT-PCR) was used to verify mRNA expression in zebrafish. Zebrafish were from the AB line (Eugene, OR; kindly provided by Dr. A. Fritz, Biology Department, Emory University). Embryos were raised at 28.5 °C, and maintained according to standard procedures in 14:10 light-dark cyclic lighting [47]. RNA from whole zebrafish larvae (96 h after fertilization) was isolated using a Trizol kit (Invitrogen). We used a Qiagen OneStep RT-PCR kit, employing a 50 µl final volume, and performed the reverse transcriptase (RT)-PCR amplifications according to the manufacturer's instructions. Gene-specific primers (Table 2) were used at final concentrations of 0.5 µM, and Mg²⁺ was at 2.5 mM. Total RNA (100 ng) was heated for 3 min at 95 °C and rapidly cooled to 4 °C to reduce secondary structure, and then 1 µg RNA was added to the reaction mix. The reverse transcription reaction was conducted at 50 °C for 30 min. Thermocycling included a single initial heat inactivation and denaturation incubation at 95 °C for 15 min, followed by 40 cycles of 94 °C for 45 s, 63 °C for 30 s, and 72 °C for 2 min. The final incubation at 72 °C was for 10 min to allow extension of partially completed PCR products. The sample was held at 4 °C until it was removed from the cycler. RT-PCR products were analyzed by 1% agarose gel electrophoresis, and stained with 1X SYBR green in water. Gel images were captured with a BioRad Gel Doc 1000 system, and cropped in Photoshop version 6.0 without any image enhancements. The PCR products were subcloned into pCR4-TOPO with a TOPO TA Cloning Kit for Sequencing (Invitrogen).

Reverse transcription and touchdown thermocycling

RNA, prepared as described above, from larval and adult zebrafish was employed, typically 100 ng per reaction. A Qiagen OneStep RT-PCR kit was used as recommended by the manufacturer. In a 50 µl reaction, 0.5 µM primer concentrations were used. The RT incubation was at 50 °C for 30 min. The RT was heat inactivated by incubating at 95 °C for 15 min. PCR was conducted using the touchdown approach. The reaction tubes were subjected to 40 cycles of 94 °C for 15 s to denature the DNA; 69 °C (decrementing the annealing temperature by 1 °C per cycle over the first 7 cycles to 63 °C) for 30 s, and an elongation step of 72 °C for 2.0 min. At the conclusion of cycling, the samples were incubated at 72 °C for 10 min as a final elongation step, and the samples were held at 4 °C until they were collected for analysis by gel electrophoresis or for subcloning and sequence analysis.

In situ hybridization

Zebrafish of the Tuebingen (Tue) strain or the albino (*albb4*) strain were maintained and bred at 28.5 °C, on a 14:10 light/dark cycle, in aquatic housing units in monitored recirculating system water. All procedures involving animals were approved by the University of Idaho Animal Care and Use Committee and conformed to the ARVO statement on the use of animals in research. Tissue from embryonic, larval, and adult zebrafish was processed for in situ hybridization as previously described [7]. In brief, adult zebrafish were anesthetized in 0.2% MS-222 (Sigma, St. Louis, MO), and decapitated. Eyes were enucleated, and then corneas were perforated and lenses removed, and eyes were placed in phosphate-buffered (pH 7.4) 4% paraformaldehyde (PFA), 5% sucrose. Larval zebrafish (99 and 155 hpf) were anesthetized in 0.02% MS-222, then immersed whole into PFA. Embryos (74 hpf) were immersed whole into PFA. All tissues were fixed for one hour, and then washed in phosphate-buffered 5% sucrose, followed by sequential washes with increasing concentrations of sucrose. Tissues were cryoprotected overnight at 4 °C in phosphate-buffered 20% sucrose, then embedded and frozen in a 2:1 solution of phosphate-buffered 20% sucrose/OCT (optimal cutting temperature embedding medium; Sakura Finetek USA, Torrance CA) and then sectioned at 3 µm.

In situ hybridizations were conducted as previously described [7]. Gene 1 and Gene 2-specific cDNAs, corresponding to PCR products amplified from zebrafish RNA with gene-specific primers (Table 1) were cloned into pCR4-TOPO-TA (Invitrogen). The plasmid with Gene 1 insert was linearized by digestion with *SpeI* or *NotI* and transcribed with T7 or T3 to generate sense and antisense (respectively) digoxigenin-labeled cRNA probes, using components of the Genius Kit (Roche). The plasmid with Gene 2 insert was linearized by digestion with *NotI* or *SpeI* and transcribed with T3 or T7 to generate sense and antisense (respectively) probes.

Sections were rehydrated and permeabilized with 10 µg/ml proteinase K for 10 min. This was followed by immersion in 0.26% acetic anhydride. Sections were then dehydrated and hybridized overnight at 56 °C. with sense or antisense Gene 1- or Gene 2-specific probes in a hybridization solution containing 50% formamide. Sections were treated with RNase A and were then incubated overnight at room temperature with anti-dig antibody conjugated to alkaline phosphatase. Hybridization was visualized with the alkaline phosphatase substrates NBT/BCIP; sections were viewed on a Leica DMR compound microscope using Nomarski optics, and were photographed with a Spot Digital camera (Diagnostic Instruments Inc., Sterling Heights, MI). Images were arranged in Photoshop CS (Adobe Systems Inc., San Jose, CA).

RESULTS

IRBP genes of tetrapods

A single IRBP locus was identified in the genomes of human, chimpanzee, dog, cow, mouse, rat, opossum, chicken, and frog. The exon-intron structures of these IRBP orthologs are mapped in Figure 1. The mammalian IRBP gene orthologs have exon and intron lengths very similar among the seven species. The exons of the chicken and frog genes are similar in length to those of the mammalian orthologs, while the chicken and frog introns are somewhat longer than their eutherian mammal counterparts. The same number of introns were found in all species, with the possible exception of the domestic dog IRBP gene, which may have an additional 3' untranslated exon (not illustrated). Gene structure of IRBP in the tetrapods appears to be highly conserved. Especially noted was the close similarity of chicken and *Xenopus tropicalis* IRBP genes.

All of the IRBP genes of tetrapods encode a single polypeptide consisting of four homologous Repeats (also referred to as modules), with each Repeat consisting of about 300 amino acids.

They all match well to the conserved domain database consensus, pfam02692.11, with bit scores ranging from about 200 to 500, corresponding to E-values of about 10^{-53} to 10^{-153} , where a bit score represents the summed information content. The bit score is derived from the raw alignment score in which the statistical properties of the scoring system are taken into account, and the bit score can be used to compare scores across different tests and types of alignments. A raw alignment score is the sum of the identity and mismatch score at each point in the sequence over a range from which the sum of gap penalties are subtracted. An identity or a substitution score is obtained from the specified weight matrix (usually BLOSUM62 for amino acid alignments).

All tetrapod IRBP orthologs have the same placement of the introns, with exon 1 encoding the first three Repeats and the first quarter of the fourth Repeat, and exons 2–4 encoding the remainder of the fourth Repeat (Figure 1). Exons 1, 2, and 3 are virtually identical in length among orthologs. Exon 4 contains a 3' untranslated region (UTR), and this 3' UTR varies widely in length across species. In general, it is difficult to define the transcription termination site in any eukaryotic gene, strictly based on consensus sequences. However, in well-studied IRBP genes [48], there are multiple transcription termination points as indicated by multiple bands on northern blots, variation in the locations of poly(A) tails in cDNA clones, multiple polyadenylation signals, and multiple lengths of the 3' UTRs in sequenced IRBP ESTs containing a poly(A) tract [48].

Figure 2 shows a multisequence alignment near each intron Donor and Acceptor site for the tetrapod orthologs. All the introns contain invariant GT and AG dinucleotides at the beginning and end of the intron. Most nearby nucleotides closely match the consensus donor or acceptor motif. The lariat sequence (not illustrated) matches the consensus in about half the introns within 50 nucleotides of the acceptor site, and all have several adenine bases, which may function in the absence of a closely matching lariat sequence. The position of each intron (relative to the coding sequence of adjacent exons) is invariant among the tetrapods.

There are several whole genome sequencing projects underway in other mammalian species, but we did not detect any matches to human IRBP in them. The African elephant (*Loxodonta africana*) and Rhesus monkey (*Macaca mulatta*) both contained an IRBP ortholog, but the sequence assemblies are incomplete, missing small parts of the IRBP gene. The IRBP genes from these species are not further considered in this report.

IRBP genes of teleosts

The genomes of four species of teleost fish have been nearly or completely sequenced. Fugu (*Takifugu rubripes*) and zebrafish (*Danio rerio*) genomes are now effectively complete, and those of pufferfish (*Tetraodon nigroviridis*) and medaka (*Oryzias latipes*) are close to completion. IRBP gene orthologs were found in each of these four genomes (Figure 3). Remarkably, in each locus from zebrafish and fugu, there were two predicted genes instead of the anticipated single IRBP gene. Each of the two genes exhibited strong sequence similarity at the amino acid level to the human IRBP sequence, indicating that both of the predicted genes were IRBP genes. The location, orientation, and intron-exon structures were similar in zebrafish and fugu. Two genes were also predicted for the IRBP locus in pufferfish; however, while the second of the genes had a very similar exon-intron pattern to the second of the IRBP genes in fugu and zebrafish, the first gene differed markedly in size. In pufferfish (Figure 3B), this first gene was much smaller and appeared to be a remnant of the first gene, at about one-fifth the size of the corresponding predicted genes from zebrafish and fugu. Despite the small size, it was unmistakable that this gene fragment encoded IRBP amino acid sequence. The medaka IRBP gene locus was different from the loci in zebrafish, fugu, and pufferfish. The medaka locus contained only a single gene corresponding to the second gene of the zebrafish and fugu loci.

The two genes predicted in each IRBP locus of fugu and zebrafish differed from the “classical” IRBP gene structure described above for tetrapods. The predicted loci were similar in structure, with the two genes separated by a short (1–2 kb) intergenic spacer. The first gene (Gene 1 hereafter) consisted of a single long exon, which encoded a protein of about 900 amino acids. The second gene (Gene 2 hereafter) had a structure virtually identical to the IRBP gene reported previously by Rajendran and coworkers [27] in zebrafish. Gene 2 in fugu and zebrafish had four exons and three introns, reminiscent of the mammalian, bird, and amphibian IRBP gene structure. The chief difference between the zebrafish and fugu Gene 2 structure and the IRBP gene of tetrapods, was that Exon 1 of the teleosts encodes just one full Repeat and a small part of a second Repeat, while Exon 1 of tetrapods encodes the first three Repeats and a small part of the fourth Repeat. The remaining exons (exons 2–4) each encode a small part of the fourth Repeat in tetrapods, and in teleosts exons 2–4 each encode an orthologous segment of the second Repeat. The positions of the introns in Gene 2 of teleosts fall almost exactly in the same positions as those in the tetrapod IRBP gene (Figure 4). The sole difference in our predicted structure of Gene 2 (Figure 3A) and the gene structure determined by Rajendran and coworkers [27] was a small exon preceding the previously established Exon 1. The small extra exon was predicted by the AUGUSTUS software but appears to be missing from the full-length cDNA sequence previously published [27]. The cDNA clone for the earlier study was obtained from zebrafish retina RNA at an adult stage [27]. While it might be an artifact of the AUGUSTUS program, it is worth noting that perhaps under some unusual circumstances, or during development when IRBP is expressed in multiple cell types (including PhRs and RPE cells [7]), this predicted upstream site might serve as a second promoter, offering potential for differential splicing or expression of the same gene in different tissues and at different times.

The predicted amino acid sequences of the teleost IRBP locus showed extensive similarity to the human IRBP amino acid sequence as illustrated in dot matrix comparisons (Figure 5A–C). These comparisons (Figure 5A) revealed five distinct diagonals looking down any column. For example, between positions 200 and 250 on the x-axis. These five diagonals indicate the presence of five Repeats in the zebrafish locus. Subsequent analyses suggest that the five Repeats are divided between two genes with Gene 1 containing the first three Repeats and Gene 2 the remaining two Repeats. A comparison of the amino acid sequence of zebrafish IRBP Gene 1 to human IRBP revealed a major diagonal (Figure 5B), indicating nearly continuous similarity over the entire 900 amino acids of the Gene 1 sequence. Every 300 amino acids an additional diagonal was found, suggesting three 300-amino acid long Repeats. The zebrafish Repeats corresponded to Repeats 1, 2, and 3 from the human sequence, with the greatest similarity to the orthologous Repeat (i.e., Repeat 1 of zebrafish was most similar to Repeat 1 of human, and revealed lesser similarities to the other three human Repeats). No Repeat 4 sequence was detected in zebrafish Gene 1. Similar studies were performed with Gene 1 from fugu and the same patterns and findings were obtained (data not shown). Comparison of the Gene 2 amino acid sequence from zebrafish to the human IRBP amino acid sequence also revealed a series of diagonals (Figure 5C). Along any column, every 300 amino acids another diagonal was found, suggesting sequence similarity to each of the four human Repeats. A major diagonal starting in the upper left corner demonstrated that Gene 2 in zebrafish begins with a Repeat 1-like motif. This diagonal line, though segmented, extends past the end of human Repeat 1 but the diagonal is heavily interrupted beyond this Repeat. Following Repeat 1 of the zebrafish, there was only one more Repeat, corresponding approximately to positions 300 to 600. Therefore, Gene 2 has only two Repeats. Next, it was apparent that the second Repeat in zebrafish Gene 2 was most similar to Repeat 4 of the human amino acid sequence. This is illustrated by the series of four diagonals vertically arranged on the right half of the dot matrix plot. The bottom diagonal was continuous, where the three other diagonals above it had more gaps and discontinuities, showing the greatest similarity to human Repeat 4. Similar results comparing Gene 2 from tetraodon, medaka, and fugu demonstrated that Gene 2 of these fish

had the same two-Repeat structure corresponding to human Repeats 1 and 4 (dot plots not shown).

Teleost genomes therefore contain IRBP loci with one or two IRBP genes. Gene 1 from zebrafish and fugu has three Repeats corresponding to Repeats 1, 2, and 3 of the human IRBP gene. Gene 1 in pufferfish contains a short single fragment of Repeat 3 (Figure 3B), while the medaka IRBP locus does not include a Gene 1 ortholog. Gene 2 from zebrafish, fugu, medaka, and pufferfish had two Repeats corresponding to human Repeats 1 and 4.

A few other teleost fish have been subjected to partial EST profiling, and we detected sequence similarities among cDNAs and ESTs from stickleback, goldfish, and fathead minnow. Dotplots comparing a consensus stickleback cDNA to the zebrafish gene locus illustrate evidence for a stickleback two-Repeat protein of the Gene 2-type, matching better to Gene 2 than Gene 1 (data not shown). The percent identity of the stickleback cDNA and Gene 2 from zebrafish is 56% and to Gene 1 is 38% at the nucleotide level. These ESTs and cDNAs illustrate the general conservation of IRBP sequences at the mRNA level. In all three species, we only detected evidence for a single kind of mRNA sequence (which was derived solely from Gene 2), and thus, only a single expressed gene, suggesting that if Gene 1 were present in these genomes, it must be poorly expressed in the pooled tissues from which RNA had been obtained.

To establish an evolutionary relationship between Gene 1 and Gene 2, we performed pairwise comparisons between the indicated two versions of Repeat 1, the only Repeat that occurs in both genes in zebrafish and fugu. There was a relative constancy in the identities and scores regardless of the origin of either Repeat 1 homolog, whether pairs of orthologs or pairs of paralogs, with the ortholog pairs more similar than the paralog pairs. This difference suggested that Genes 1 and 2 diverged before zebrafish and fugu diverged (Table 3). A phylogeny of Repeat 1 is shown in Figure 6 to indicate relative evolutionary distances. This phylogeny suggests that Gene 1 and Gene 2 were created early and simultaneously in teleost evolution, well before fugu and zebrafish last shared a common ancestor.

We compared intron locations across the IRBP orthologs. Consensus maps of the tetrapod and teleost fish IRBP genes at each orthologous splice site boundary, predicted by computer and by hand, are provided in Figure 2 and Figure 4, respectively. The consensus splice site sequences are taken from Zhang [49]. The donor and acceptor sites in IRBP closely match the consensus splice sites, with invariant GT...AG sequences at the beginning and end of each intron. There is little variation in the position of introns in the IRBP gene. However, within the intron sequences, except near the splice sites, intronic sequences are not conserved. This was shown by dot matrix comparison of fugu and zebrafish (representing the teleosts) and among several of the tetrapods (data not shown).

We performed multiple sequence alignments of the predicted amino acid sequences of teleost fish IRBP Gene 1 and Gene 2 at the amino acid sequence level (Figure 7). First, we aligned the Gene 1 amino acid sequences from zebrafish and fugu (Figure 7A). Among the 918 aligned positions, only two gaps were inserted totaling 5 amino acids. The sequences are overall about 58% identical. Next, we compared the Repeat 1 amino acid sequence from Gene 2 of several fish. The sequences of six species are shown in Figure 7B, including fugu, goldfish, medaka, stickleback, tetraodon, and zebrafish. Numerous blocks of identical and strongly conserved amino acids are found throughout the entire length of the sequence. There are no long deletions or insertions in any of the aligned sequences, suggesting that no domains have been gained or lost. The Repeat structure in IRBP contains two domains, designated A and B, and both are conserved among all the Repeats in the teleost fish. Potential glycosylation sites are marked and have a conserved location about 200 amino acids from the N-terminal end of Repeat 1, which is similar to the tetrapods. One of the potential hyaluronan binding sites is shared with

the mammals, at about positions 220–230 of Repeat 1. This is conserved in all 6 fish (Figure 7A,B).

Nonvertebrate chordate IRBP gene searches

We were not able to identify an IRBP gene ortholog in *Ciona intestinalis* or *Ciona savignyi*, two urochordates that lack complex eyes. The most current versions of the genome assemblies of these species were screened for matching sequence on July 8, 2005. Because all vertebrates with “camera-like” eyes have IRBP, and because these two chordates lack IRBP, we suggest that IRBP arose after the divergence of the urochordates from the vertebrates. It will be interesting and useful to search for an ortholog of IRBP in the cephalochordates that have eyes and ciliary photoreceptors; this will be possible when the amphioxus whole genome sequencing project is complete.

Expression of two IRBP genes in zebrafish

To determine whether both Gene 1 and Gene 2 of zebrafish were transcribed into stable mRNAs, we performed gene-specific RT-PCRs. A primer pair from Gene 1 was designed to amplify a 1529 bp band, if an mRNA was transcribed from Gene 1. Because this gene lacks any introns, it was not possible to design a primer pair that spans an intron. However, Gene 2 has three introns, and a primer pair was designed that spanned Introns B and C. If Gene 2 was transcribed and spliced, a band of 532 bp was expected. In both cases, bands of expected size were amplified from 96 h larvae and adult whole eye RNA with each primer set (Figure 8A). In the absence of the reverse transcriptase, no PCR products were detected, suggesting that there was no genomic DNA contamination in the RNA (data not shown). In Gene 2 RT-PCRs, splicing was shown to occur as the RT-PCR product size agreed closely with the expected size based on the removal of the intron from the processed mRNA. These data indicate that both genes in zebrafish are transcriptionally active.

The gene structure of the zebrafish IRBP locus is consistent with the hypothetical formation of a single individual transcript that would include both genes. To determine if this is possible, we performed RT-PCR using a forward primer about 500 bp upstream from the end of the zebrafish Gene 1 and a reverse primer located about 600 bp downstream from the beginning of zebrafish Gene 2. This experiment resulted in amplification of a product of a size lacking the intergenic spacer (Figure 8B). This prominent band was about 1100 bp (expected size of 1098 bp). A less prominent co-migrating band was also amplified from larvae (Figure 8B).

The amplified RT-PCR product from 96 h larvae was cloned and sequenced. It was compared to the genomic sequence of the AB strain of zebrafish, which was obtained by cloning PCR amplified genomic DNA from three AB fish. In sequencing the three independent genomic clones of the AB strain, all three gave the identical sequence. The resulting RT-PCR sequence contained, in order, a priming site for primer F1 (Table 2), about 550 nt of Gene 1, about 50 nt of 3' UTR sequence not homologous to Gene 1, about 500 nt of Gene 2, and a priming site for R3 (Table 2). The overall length and sequence of RT-PCR amplified product is most consistent with an RNA transcript originating in Gene 1, transcribing through the intergenic region into Gene 2, followed by splicing to remove the intergenic spacer. The 50 bp nonhomologous sequence reflects a 50 bp sequence inversion in the Gene 1 3' UTR, as this inverted sequence was detected when the opposite strands are aligned. It is not clear whether the inversion represents a rare RNA processing error, a cloning or reverse transcription error, or a rare genomic sequence change within the AB strain of zebrafish. Other than the loss of the intergenic spacer and the short sequence inversion, the RT-PCR and genomic sequences were identical.

The loss of the intergenic spacer from the primary transcript suggests that the spacer may, under some circumstances, function as an intron, when transcription runs beyond the end of Gene 1. This appears to represent aberrant and rare transcription run-through and splicing in the zebrafish IRBP gene locus.

The presence and expression of two IRBP genes in the zebrafish and fugu genomes, but not in other genomes, suggests that in the former species, the two paralogous IRBP genes may have assumed divergent roles [40,50]. Similar neo- and sub-functionalization processes for other duplicated genes have included the emergence of differential expression patterns [39,40]. We performed *in situ* hybridization with Gene 1- and Gene 2-specific cRNA probes to determine if this was the case for zebrafish IRBP. Consistent with previous findings [7], Gene 2 is expressed in photoreceptors and the RPE, in 74 hpf embryonic, 99 and 155 hpf larval, and adult retinas (Figure 9). In addition, very weak and sporadic expression of Gene 2 was observed [7] in a minor subpopulation of cells in the inner nuclear layer (INL), Figure 9. In contrast, Gene 1 was expressed in a slightly larger proportion of cells residing in the INL, with positive signals weakly detected in embryonic retinas and stronger signals in larval and adult retinas (Figure 9). Expression of Gene 1 was not localized to photoreceptors or to the RPE; however, we occasionally observed expression of Gene 1 in a subpopulation of cells in the ganglion cell layer (data not shown). Both Gene 1 and Gene 2 were expressed in the pineal organ in embryos and larvae, although positive signals were barely detectable for Gene 1 (Figure 9). We did not evaluate IRBP expression in adult pineal organs. The use of sense probes corresponding to Gene 1 or Gene 2 resulted in no labeling (data not shown). These data collectively demonstrate differential spatiotemporal expression of Gene 1 and Gene 2.

DISCUSSION

The structure of the tetrapod IRBP gene

The tetrapod IRBP gene structure is now well established and consistent among multiple taxa. Tetrapod IRBP has four exons and three introns, with a large first exon encoding the first three Repeats and the beginning of the fourth (and last) Repeat. This conclusion is supported by the close sequence similarities at donor and acceptor splice sites (Figure 2) and the close amino acid sequence similarities among the different species in each exon. The IRBP gene is highly conserved among the tetrapods, consistent with an important function in the visual system.

There are, however, exceptions to this conserved structure. For example, our analyses predicted that the IRBP gene of the domestic dog may possess an extra intron at the end of the gene. This prediction awaits proof through sequence analysis of the expressed dog IRBP mRNA. An additional 246 nt was also identified in the fourth exon of an inbred jungle fowl IRBP mRNA; this in-frame sequence was not present in the domestic chicken genomic sequence, and appears to be a duplication of a sequence found in the second intron [51]. It is worth mentioning that the opossum and chicken possess long introns. Run-through transcription and multiple polyadenylation sites in the mouse gene, and a large insertion into the 3' UTR of the bovine gene near the 3' end of the gene also reflect variability in the structure of the tetrapod IRBP gene at the 3' end. It does not seem likely though that the Repeat structure of IRBP makes this gene susceptible to run-through or run-on transcription, which seems common among most eukaryotic multicellular organisms.

The structure of the IRBP gene locus in teleosts and models of teleost IRBP evolution

The teleost IRBP gene locus varies in structure among the four teleost fish studied here (Figure 3), and has revealed an intriguing evolutionary history of the IRBP gene. There are clear signs of a gene duplication, resulting in a head-to-tail two-gene locus containing two different (but related) IRBP genes. It appears that one of the genes (Gene 1) is in some cases undergoing

gene loss or has been relegated to pseudogene status (as observed in medaka and pufferfish) and in other cases potentially has undergone neo- or sub-functionalization (in zebrafish and perhaps in fugu). The second gene in the locus (Gene 2) is uniformly retained in structure in all four species. The two genes contain remarkably different gene structures: Gene 1 consists of a single long exon encoding a 900 amino acid long protein encoding Repeats 1, 2, and 3; Gene 2 is a more typical gene consisting of four exons that encode a polypeptide consisting of two full Repeats, Repeats 1 and 4.

We suggest that the two IRBP genes arose early in the evolution of teleosts, at approximately the time of the whole genome duplication (WGD) that occurred coincident with or just prior to the radiation of the teleosts [39,40,42,50]. The duplication occurred early in the evolution of the teleosts as the two-gene locus is found in the distantly related zebrafish and fugu. This hypothesis is further supported by the presence of Gene 2, preceded by a remnant of Gene 1 in pufferfish, and by the closer sequence similarities of orthologs of Repeat 1 than of paralogs of Repeat 1. We predict that many other teleost fish will have two IRBP genes of similar exon and intron structures in a single locus (Figure 3). The existence of a three-Repeat Gene 1 in fugu, but a Gene 1 remnant in the related pufferfish, and the absence of Gene 1 entirely in the more distantly related medaka, also suggests that the loss or pseudogenation of Gene 1 has occurred as multiple independent events. Perhaps the study of comparative visual requirements in these species may reveal hints as to the function of Gene 1.

It is remarkable that the duplicated IRBP genes have only one Repeat - Repeat 1 - in common. Repeats 2 and 3 are lacking from Gene 2, and Repeat 4 is not present in Gene 1. Pairwise comparisons between any two homologs of Repeat 1 (Table 3), indicated that the Repeat 1 homologs diverged from one another at about the same time. It is tempting to speculate that Repeat 1 may correspond most closely to a putative ancestral single-Repeat IRBP, though Repeat 4 might be the original, as indicated by the presence of introns. The predicted timing of Repeat 1 divergence is consistent with the already known radiation of the teleost fish about 350 Mya. The more important consideration is that roughly the same results were found when comparing the paralogs within one species (Table 3). These comparisons, one in zebrafish and one in fugu, suggest that the two IRBP genes (Gene 1 and Gene 2) arose approximately coincident with the radiation of the teleosts and the proposed whole genome duplication, and the two-gene locus in the euteleosts arose early in the ancestry of these fish. We hypothesize that this event, the creation of the two-gene IRBP locus, may serve as a marker of the euteleosts. A corollary of this hypothesis is that more primitive fish, including coelacanths, lungfish, bow fin, and gar, are predicted to have a single gene IRBP locus, with a four-Repeat IRBP gene structure characteristic of the tetrapods. This hypothesis will be tested when whole genome sequencing projects in these and related species are completed.

Mechanisms causing the tandem IRBP gene duplication

Several evolutionary models are consistent with the observed gene structures of the teleost IRBP loci. Here we describe what we believe to be the two most parsimonious models (Figure 10 and Figure 11). For both models, the tetrapods, teleosts, and cartilaginous fish are considered to have a common ancestor having a full four-repeat, four-exon IRBP gene. The evolutionary quadruplication of the even more ancestral single Repeat IRBP gene is still anticipated to occur at about the time when vertebrates first arose during the Silurian period. The other essential requirements are that a model must result in a single locus bearing two IRBP genes in head to tail orientation, with no other IRBP loci in the fish genome.

In the first model (Figure 10), the teleost IRBP two-gene locus is proposed to have arisen as a direct consequence of the teleost WGD, because they (the WGD and the two-gene locus) appeared coincidentally in evolutionary time. The WGD generated two complete four-Repeat IRBP genes, one on each of the duplicated whole chromatids. These duplicates are proposed

to have diverged as illustrated in Figure 10, steps 2A, 2B, and 2C, in which each one of the genes loses one or more Repeats and in which there are alterations to the cis-elements in the promoters. Following divergence, a reduction in size of the tetraploid ancestral teleost genome began to occur. The reduction in size is proposed to have occurred by the unequal crossing-over of slightly diverged chromatids and the loss of the chromatid that lacks an IRBP gene (presumably by natural selection). The sequence divergence that had occurred on the two IRBP loci may have contributed to a crossover event that yielded one new chromatid with two IRBP genes, one gene with three Repeats and a second gene, which has only two Repeats, in head to tail orientation with the first. This is a mechanism that preserves synteny, and it results in only minor revisions to the current model (Figure 12) of IRBP evolution [26].

A second model for the origin of the two IRBP tandem IRBP genes, is a single-gene tandem duplication (Figure 11). This mechanism involves the simple misalignment of two chromatids that are undergoing recombination (Figure 11, Step 1). In Step 1, the unequal cross-over is external to the IRBP gene. In Steps 2 and 3, each of the tandem genes must undergo the loss of internal segments of the gene, with Gene 1 losing Repeat 4-encoding DNA, and Gene 2 losing DNA that encodes Repeats 2 and 3. These two steps differ little from similar steps illustrated in the previous model shown in Figure 10. The single-gene tandem duplication model also preserves synteny. This mechanism is well established in many other gene families, and is considered to be the mechanism for the generation of tandem repeats of some of the cone opsin genes [52,53]. Our estimated evolutionary timing of the original gene duplication event predicts that either some teleosts would have two additional IRBP genes in their genomes (if the WGD occurred after the single-gene tandem duplication), or one additional IRBP gene (if the WGD occurred first). We have not found evidence of such additional IRBP genes; however, we have only examined the genomes of four teleosts, and there is evidence that gene loss has eliminated a large proportion of duplicate genes in teleosts [54,55].

It is hypothetically possible to compare the odds of the two models, the tandem duplication (Figure 11) and WGD/compaction (Figure 10) models. Assuming that the probability of retaining a gene duplicate is 24% after the WGD [56], then the two models in Figures 10 and 11 are about equally likely, with the WGD model slightly favored by about 1.6 fold. The complete genome sequences of several particular teleosts should lead to a resolution of which model is responsible for the duplication of the IRBP gene.

Several alternative mechanisms are reasonable to propose, but a detailed discussion is beyond the scope of this paper. We wish to emphasize that, since the IRBP gene duplication cannot be observed directly, we can only make inferences based on limited data and probability. Thus, all these models are to some degree speculative, but they serve as useful hypotheses to test with more sequence data. We hope to build more accurate models and to better distinguish among these models once we obtain whole genome sequence data from several more fish species.

Potential neo- and sub-functionalization of IRBP Gene 1 and Gene 2 in zebrafish

Expression patterns and regulation of IRBP in zebrafish have been the subject of considerable investigation. Expression of zebrafish IRBP mRNA is strongly diurnal and has been shown to be circadian in the eye [27,57] and in the pineal [58]. IRBP mRNA is expressed at high levels in light and at low levels in dark, out of phase with CLOCK [57]. IRBP mRNA transcription in the pineal organ is dependent on Otx5 [58]. In addition, zebrafish IRBP mRNA is expressed by the RPE as well as by photoreceptors. This pattern is distinct, as the IRBP of all other vertebrates examined is photoreceptor-specific [7]. Our data introduce an important further consideration in the study of zebrafish IRBP. We provide evidence that two IRBP genes are not only present, but are expressed, in the zebrafish eye. The RT-PCR and in situ hybridization results lend support to the putative neofunctionalization or sub-function partitioning [40,50]

of the two IRBP genes. The differential cell-type specific expression from each gene, and the consequent division of the four-Repeat protein into Repeats 1, 2, and 3 in Gene 1 and Repeats 1 and 4 in Gene 2, now must be considered in re-evaluating expression and regulation of IRBP. In addition, selective gene-specific knockdown strategies in the zebrafish may allow the elucidation of the functions of each IRBP gene.

Gene 1 and Gene 2 are both transcribed, and Gene 2 is post-transcriptionally processed as reflected by the loss of at least one intron (Figure 8). These studies also demonstrated that a primary transcript can initiate upstream of Gene 1 and elongate through Gene 2, and is spliced to produce an mRNA lacking the intergenic spacer. It is possible that this spliced mRNA might encode five Repeats. Repeats 1–3 (from Gene 1) may be fused in frame to Repeat 1 and 4 from Gene 2 in a single polypeptide. We have not yet evaluated cell-specific expression of this long, but rare, mRNA.

It was clear from earlier experiments [7] that IRBP was expressed heavily in photoreceptor and RPE cells and pinealocytes [58]. However, Gene 1- and 2-specific expression patterns were not discriminated because Repeat 1-specific sequences are contained in both genes, and the prior probes included Repeat 1. Repeat 1 from Gene 1 exhibits about 70% identity at the nucleotide level to Repeat 1 from Gene 2 (data not shown), which may be sufficient for cross-hybridization. With the use of the gene-specific probes produced here, we tested the hypothesis that Gene 1 and Gene 2 were differentially expressed. The results clearly demonstrated that there was differential expression of the two genes, but we were surprised by the identity of the cell types that expressed Gene 1. First, there was no extensive overlap in expression, with each Gene being expressed predominantly in different cell types. Gene 2 was expressed in the previously detected pattern in both photoreceptor and RPE cells [7]. However, Gene 1 showed an unexpected and novel expression pattern, in a sub-population of cells in the INL, and occasionally in a subpopulation of cells in the ganglion cell layer. The distinct expression patterns of Gene 1 and Gene 2 are consistent with neo- and sub-functionalization of the two IRBP genes. We have not yet explored whether any of the INL cells that express Gene 1, also express Gene 2. Pursuit of the identity of the Gene 1-expressing cells is underway.

The IRBP gene locus, with two genes having independent expression patterns and therefore presumed independent functions, can now be added to the inventory of examples of neo- or sub-functionalization in light-sensitive tissues. Extra-retinal opsin (*errlo*) shares approximately 74% identity at the amino acid level with rod-specific opsin (*rh*) from the retina of the same species of teleost fish [59,60]. *Errlo* is expressed in the pineal gland but not in the retina, and *rh* is expressed in the retina but not in the pineal. *Errlo* bears introns much like the ancestral opsin gene, and appears to be the ortholog of the mammalian rhodopsin gene. The *rh* gene is intronless and thought to be a retrogene that integrated just in front of the intron-containing *errlo* gene [61,62]. Bellingham et al. [62] found that the *rh* retrogene was formed at or before the appearance of sturgeon, bichir, and gar, events that preceded the WGD of the euteleosts [41]. As suggested for *rh* and *errlo* [60], differences between IRBP Gene 1 and Gene 2 may have arisen from their differing functional roles or cell-type specific expression patterns. In future studies we will report on the promoter structure-function studies of the teleost IRBP genes. Another example of neo- or sub-functionalization in the eye is the evolution of guanylate cyclase-activating proteins (GCAPs) [63], which are known to regulate photoreceptor guanylate cyclases (GCs). Baehr and coworkers [63] found evidence of eight GCAP genes in fugu. The diversity and number of these genes, and their differential expression, suggests that these “extra” GCAP genes may have functions other than or in addition to the stimulation of GCs. Finally, teleost genomes contain multiple genes encoding some of the cone opsin genes. Zebrafish has two red (LWS/MWS), and three green (RH2) genes, with each set arranged head-to-tail on separate chromosomes [52]. The situation is similar in medaka, although medaka also has a tandem duplication of the blue (SWS2) gene, and only three copies of the RH2 gene

[53]. There is evidence for some sub-functionalization of the cone opsins in the zebrafish, as the gene copies show differential spatiotemporal expression patterns [64], and divergent absorption spectra [52]. These examples highlight the robust and broad utility to the study of gene locus structure and gene expression in the teleost fish. Partitioning of discrete tasks between two genes offers tremendous promise as a general approach to determine the role(s) of a gene with no known property or physiological activity.

The origins of IRBP

It has been known for many years that IRBP is present in vertebrates and absent from the invertebrates. It is also known that only teleost IRBP lacked the four-Repeat structure of the typical vertebrate IRBP [27]. For example, IRBP in the little skate, *Leucoraja erinacea*, [65] and the dogfish, *Squalus acanthias*, [49] is a large protein about the same size as tetrapod IRBP. This suggested [27] that the IRBP gene and protein from the elasmobranchi (sharks, skates, and rays; that is, all cartilaginous fish) have the same four-Repeat, 1200 amino acid long polypeptide and the hypothesized four-exon three-intron structure in the most recent common ancestor of all these species. That these species share common ancestors predating the whole genome duplication in the teleosts, suggests that the ancestral form of the vertebrate IRBP gene was the four Repeat structure containing four exons and three introns, similar to that seen in present-day mammals, birds, and amphibians. Proof of this hypothesized gene and protein structure in the rays and sharks awaits the completion of whole genome sequencing projects that are already underway for representative species of the cartilaginous fish [45].

The revised model of IRBP evolution may be further tested by determining the IRBP gene structures in additional taxa of the chordates. In particular we hypothesize that the non-teleost fish (cf., jawless fish, cartilaginous fish, lobe fin fish) should contain a gene structure similar to the mammalian structure with four Repeats, roughly encoding a 1200 amino acid long polypeptide (Figure 12).

Summary

The tetrapod IRBP gene structure is tightly conserved. The same number of Repeats and the same gene structure was found in mammals, birds, and amphibians. In the teleosts, the gene and locus structure vary substantially, but in a systematic pattern. In teleost fish, the number of Repeats is different from the mammals in each of the two IRBP genes. In two species, the two-Repeat gene is preceded by an intronless three-Repeat gene. We have revised a model of IRBP gene evolution (Figure 12) to be consistent with these gene structures. The absence of an IRBP gene in urochordates suggests that an original ancestral IRBP gene arose in an interval after the urochordates diverged but before the vertebrates diverged from other chordates. The gene created then was of the tetrapod 4-exon, 3-intron, 4 repeat type. Here we revise the evolutionary history of IRBP in the teleosts, but not in the history of tetrapod or elasmobranch evolution.

The zebrafish gene expression patterns support sub-function partitioning or neofunctionalization of the two teleost IRBP genes. Every teleost that has been previously examined [65] appears to express Gene 2 (based on the size of the protein), and here we showed that in zebrafish, Gene 2 is expressed in PhRs and RPE, establishing that Repeats 1 and 4 may be required for vision. Repeats 2 and 3 were expressed uniquely in INL and perhaps GCL cells and therefore may be necessary for other processes. Teleost fish are different from tetrapods and consequently may differ in their requirements for the types of IRBP Repeats. This difference may reflect different evolutionary pressures, such as genome compaction, or different physiological, environmental, or physical requirements in vision.

Acknowledgments

This study was supported by NIH R01EY012146 (DLS), K12GM000680, R01EY016470, R03EY013986, R24EY017045, and P30EY006360, the Foundation Fighting Blindness, Fight for Sight, Research to Prevent Blindness Inc., and Knights Templar of Georgia.

We thank several major sequencing centers for making whole genome sequences available for these analyses. In accordance with the wishes of these centers, we quote each requested acknowledgment below.

“The zebrafish sequence data were produced by the Zebrafish Sequencing Group at the Sanger Institute and can be obtained from D_rerio. The zebrafish draft assemblies were provided by The Wellcome Trust Sanger Institute, Cambridge, UK.”

“The *Xenopus tropicalis* EST sequence data were produced by the *Xenopus tropicalis* Sequencing Group at the Sanger Institute and can be obtained from X_tropicalis.”

“The *Tetraodon nigroviridis* V7 assembly (February 2004) was provided by Genoscope, Evry, France in collaboration with The Broad Institute, Cambridge, MA.”

We thank the Medaka Genome Sequencing Project: “The data has been provided freely by the National Institute of Genetics and the University of Tokyo for use in this publication/correspondence only.”

“The opossum sequence was made freely available by The Broad Institute, Cambridge, MA.”

“The February 2004 chicken draft sequence was produced by The Genome Sequencing Center at Washington University School of Medicine, St. Louis, MO.”

The fugu IRBP gene locus sequence “has been provided freely by the Fugu Genome Consortium for use in this publication/correspondence only.”

Stickleback and fathead minnow ESTs were obtained from GenBank and the sequences were kindly deposited by the Stanford Human Genome Center, the US EPA, and the DOE Joint Genome Institute Pimephales promelas EST project.

References

1. Bunt-Milam AH, Saari JC. Immunocytochemical localization of two retinoid-binding proteins in vertebrate retina. *J Cell Biol* 1983;97:703–12. [PubMed: 6350319]
2. Crouch RK, Hazard ES, Lind T, Wiggert B, Chader G, Corson DW. Interphotoreceptor retinoid-binding protein and alpha-tocopherol preserve the isomeric and oxidation state of retinol. *Photochem Photobiol* 1992;56:251–5. [PubMed: 1502268]
3. Gonzalez-Fernandez F. Evolution of the visual cycle: the role of retinoid-binding proteins. *J Endocrinol* 2002;175:75–88. [PubMed: 12379492]
4. Liou GI, Wang M, Matragoon S. Timing of interphotoreceptor retinoid-binding protein (IRBP) gene expression and hypomethylation in developing mouse retina. *Dev Biol* 1994;161:345–56. [PubMed: 8313988]
5. Gonzalez-Fernandez F, Van Niel E, Edmonds C, Beaver H, Nickerson JM, Garcia-Fernandez JM, Campohiaro PA, Foster RG. Differential expression of interphotoreceptor retinoid-binding protein, opsin, cellular retinaldehyde-binding protein, and basic fibroblastic growth factor. *Exp Eye Res* 1993;56:411–27. Erratum in: *Exp Eye Res* 1993 Jul;57(1):127. [PubMed: 7916695]
6. Timmers AM, Newton BR, Hauswirth WW. Synthesis and stability of retinal photoreceptor mRNAs are coordinately regulated during bovine fetal development. *Exp Eye Res* 1993;56:257–65. [PubMed: 8386100]
7. Stenkamp DL, Cunningham LL, Raymond PA, Gonzalez-Fernandez F. Novel expression pattern of interphotoreceptor retinoid-binding protein (IRBP) in the adult and developing zebrafish retina and RPE. *Mol Vis* 1998;4:26. [PubMed: 9841935]
8. Liou GI, Fei Y, Peachey NS, Matragoon S, Wei S, Blaner WS, Wang Y, Liu C, Gottesman ME, Ripps H. Early onset photoreceptor abnormalities induced by targeted disruption of the interphotoreceptor retinoid-binding protein gene. *J Neurosci* 1998;18:4511–20. [PubMed: 9614228]

9. Fong SL, Fong WB. Elements regulating the transcription of human interstitial retinoid-binding protein (IRBP) gene in cultured retinoblastoma cells. *Curr Eye Res* 1999;18:283–91. [PubMed: 10372988]
10. Furukawa T, Morrow EM, Li T, Davis FC, Cepko CL. Retinopathy and attenuated circadian entrainment in Crx-deficient mice. *Nat Genet* 1999;23:466–70. [PubMed: 10581037]
11. Saari JC, Teller DC, Crabb JW, Bredberg L. Properties of an interphotoreceptor retinoid-binding protein from bovine retina. *J Biol Chem* 1985;260:195–201. [PubMed: 2981203]
12. Lai YL, Wiggert B, Liu YP, Chader GJ. Interphotoreceptor retinol-binding proteins: possible transport vehicles between compartments of the retina. *Nature* 1982;298:848–9. [PubMed: 7202123]
13. Fong SL, Liou GI, Landers RA, Alvarez RA, Bridges CD. Purification and characterization of a retinol-binding glycoprotein synthesized and secreted by bovine neural retina. *J Biol Chem* 1984;259:6534–42. [PubMed: 6427217]
14. Chen Y, Houghton LA, Brenna JT, Noy N. Docosahexaenoic acid modulates the interactions of the interphotoreceptor retinoid-binding protein with 11-cis-retinal. *J Biol Chem* 1996;271:20507–15. [PubMed: 8702792]
15. Chen Y, Saari JC, Noy N. Interactions of all-trans-retinol and long-chain fatty acids with interphotoreceptor retinoid-binding protein. *Biochemistry* 1993;32:11311–8. [PubMed: 8218196]
16. Ho MT, Massey JB, Pownall HJ, Anderson RE, Hollyfield JG. Mechanism of vitamin A movement between rod outer segments, interphotoreceptor retinoid-binding protein, and liposomes. *J Biol Chem* 1989;264:928–35. [PubMed: 2910871]
17. Nickerson JM, Li GR, Lin ZY, Takizawa N, Si JS, Gross EA. Structure-function relationships in the four repeats of human interphotoreceptor retinoid-binding protein (IRBP). *Mol Vis* 1998;4:33. [PubMed: 9873071]
18. Lin ZY, Li GR, Takizawa N, Si JS, Gross EA, Richardson K, Nickerson JM. Structure-function relationships in interphotoreceptor retinoid-binding protein (IRBP). *Mol Vis* 1997;3:17. [PubMed: 9479008]
19. Lin ZY, Si JS, Nickerson JM. Biochemical and biophysical properties of recombinant human interphotoreceptor retinoid binding protein. *Invest Ophthalmol Vis Sci* 1994;35:3599–612. [PubMed: 8088950]
20. Baer CA, Retief JD, Van Niel E, Braiman MS, Gonzalez-Fernandez F. Soluble expression in *E. coli* of a functional interphotoreceptor retinoid-binding protein module fused to thioredoxin: correlation of vitamin A binding regions with conserved domains of C-terminal processing proteases. *Exp Eye Res* 1998;66:249–62. [PubMed: 9533851]
21. Hessler RB, Baer CA, Bukelman A, Kittredge KL, Gonzalez-Fernandez F. Interphotoreceptor retinoid-binding protein (IRBP): expression in the adult and developing *Xenopus* retina. *J Comp Neurol* 1996;367:329–41. [PubMed: 8698895]
22. Baer CA, Kittredge KL, Klinger AL, Briercheck DM, Braiman MS, Gonzalez-Fernandez F. Expression and characterization of the fourth repeat of *Xenopus* interphotoreceptor retinoid-binding protein in *E. coli*. *Curr Eye Res* 1994;13:391–400. [PubMed: 7924403]
23. Gonzalez-Fernandez F, Kittredge KL, Rayborn ME, Hollyfield JG, Landers RA, Saha M, Grainger RM. Interphotoreceptor retinoid-binding protein (IRBP), a major 124 kDa glycoprotein in the interphotoreceptor matrix of *Xenopus laevis*. Characterization, molecular cloning and biosynthesis. *J Cell Sci* 1993;105 (Pt 1):7–21. [PubMed: 8360278]
24. Lamb TD, Pugh EN Jr. Dark adaptation and the retinoid cycle of vision. *Prog Retin Eye Res* 2004;23:307–80. [PubMed: 15177205]
25. Ripps H, Peachey NS, Xu X, Nozell SE, Smith SB, Liou GI. The rhodopsin cycle is preserved in IRBP “knockout” mice despite abnormalities in retinal structure and function. *Vis Neurosci* 2000 Jan–Feb;17:97–105. [PubMed: 10750831]
26. Borst DE, Redmond TM, Elser JE, Gonda MA, Wiggert B, Chader GJ, Nickerson JM. Interphotoreceptor retinoid-binding protein. Gene characterization, protein repeat structure, and its evolution. *J Biol Chem* 1989;264:1115–23. [PubMed: 2910846]
27. Rajendran RR, Van Niel EE, Stenkamp DL, Cunningham LL, Raymond PA, Gonzalez-Fernandez F. Zebrafish interphotoreceptor retinoid-binding protein: differential circadian expression among cone subtypes. *J Exp Biol* 1996;199:2775–87. [PubMed: 9110959]

28. Poux C, Douzery EJ. Primate phylogeny, evolutionary rate variations, and divergence times: a contribution from the nuclear gene IRBP. *Am J Phys Anthropol* 2004;124:1–16. [PubMed: 15085543]
29. Stanhope MJ, Smith MR, Waddell VG, Porter CA, Shivji MS, Goodman M. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J Mol Evol* 1996;43:83–92. [PubMed: 8660440]
30. Loew A, Gonzalez-Fernandez F. Crystal structure of the functional unit of interphotoreceptor retinoid binding protein. *Structure* 2002;10:43–9. [PubMed: 11796109]
31. Loew A, Baer C, Gonzalez-Fernandez F. The functional unit of interphotoreceptor retinoid-binding protein (IRBP)—purification, characterization and preliminary crystallographic analysis. *Exp Eye Res* 2001;73:257–64. [PubMed: 11446776]
32. Engel CK, Mathieu M, Zeelen JP, Hiltunen JK, Wierenga RK. Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket. *EMBO J* 1996;15:5135–45. [PubMed: 8895557]
33. Modis Y, Filippula SA, Novikov DK, Norledge B, Hiltunen JK, Wierenga RK. The crystal structure of dienoyl-CoA isomerase at 1.5 Å resolution reveals the importance of aspartate and glutamate sidechains for catalysis. *Structure* 1998;6:957–70. [PubMed: 9739087]
34. Benning MM, Taylor KL, Liu R-Q, Yang G, Xiang H, Wesenberg G, Dunaway-Mariano D, Holden HM. Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: an enzyme catalyst generated via adaptive mutation. *Biochemistry* 1996;35:8103–9. [PubMed: 8679561]
35. Liao DI, Qian J, Chisholm DA, Jordan DB, Diner BA. Crystal structures of the photosystem II D1 C-terminal processing protease. *Nat Struct Biol* 2000;7:749–53. [PubMed: 10966643]
36. Silber KR, Keiler KC, Sauer RT. Tsp: a tail-specific protease that selectively degrades proteins with nonpolar C termini. *Proc Natl Acad Sci U S A* 1992;89:295–9. [PubMed: 1729701]
37. Gross EA, Li GR, Lin ZY, Ruuska SE, Boatright JH, Mian IS, Nickerson JM. Prediction of structural and functional relationships of Repeat 1 of human interphotoreceptor retinoid-binding protein (IRBP) with other proteins. *Mol Vis* 2000;6:30–9. [PubMed: 10756179]
38. Loew A, Gonzalez-Fernandez F. X-ray structure of the second module of xenopus interphotoreceptor retinoid-binding protein. *Invest Ophthalmol Vis Sci* 2001;42:S356.
39. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* 2004;38:615–43. [PubMed: 15568988]
40. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999;151:1531–45. [PubMed: 10101175]
41. Hoegg S, Brinkmann H, Taylor JS, Meyer A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol* 2004;59:190–203. [PubMed: 15486693]
42. Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* 2004;21:1146–51. [PubMed: 15014147]
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10. [PubMed: 2231712]
44. Borst DE, Nickerson JM. The isolation of a gene encoding interphotoreceptor retinoid-binding protein. *Exp Eye Res* 1988;47:825–38. [PubMed: 3215295]
45. Venkatesh B, Tay A, Dandona N, Patil JG, Brenner S. A compact cartilaginous fish model genome. *Curr Biol* 2005;15:R82–3. [PubMed: 15694293]
46. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 2004;32:W309–12. [PubMed: 15215400]
47. Westerfield, M. A guide for the laboratory use of zebrafish (*Danio rerio*). 4. Univ. of Oregon Press; Eugene: 2000. The zebrafish book.
48. Shuler RK Jr, Gross E, He WY, Liou GI, Nickerson JM. Sequence analysis of the mouse IRBP gene and cDNA. *Curr Eye Res* 2002;24:354–67. [PubMed: 12434304]

49. Duffy M, Sun Y, Wiggert B, Duncan T, Chader GJ, Ripps H. Interphotoreceptor retinoid binding protein (IRBP) enhances rhodopsin regeneration in the experimentally detached retina. *Exp Eye Res* 1993;57:771–82. [PubMed: 8150029]
50. Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet* 2004;20:481–90. [PubMed: 15363902]
51. Stenkamp DL, Calderwood JL, Van Niel EE, Daniels LM, Gonzalez-Fernandez F. The interphotoreceptor retinoid-binding protein (IRBP) of the chicken (*Gallus gallus domesticus*). *Mol Vis* 2005;11:833–45. [PubMed: 16254552]
52. Chinen A, Hamaoka T, Yamada Y, Kawamura S. Gene duplication and spectral diversification of cone visual pigments of zebrafish. *Genetics* 2003;163:663–75. [PubMed: 12618404]
53. Matsumoto Y, Fukamachi S, Mitani H, Kawamura S. Functional characterization of visual opsin repertoire in Medaka (*Oryzias latipes*). *Gene* 2006;371:268–78. [PubMed: 16460888]
54. Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res* 2005;15:1307–14. [PubMed: 16109975]
55. Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, Huang H, Postlethwait JH, Talbot WS. A comparative map of the zebrafish genome. *Genome Res* 2000;10:1903–14. [PubMed: 11116086]
56. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 2004;431:946–57. [PubMed: 15496914]
57. Whitmore D, Foulkes NS, Strahle U, Sassone-Corsi P. Zebrafish Clock rhythmic expression reveals independent peripheral circadian oscillators. *Nat Neurosci* 1998;1:701–7. [PubMed: 10196586]
58. Gamse JT, Shen YC, Thisse C, Thisse B, Raymond PA, Halpern ME, Liang JO. *Otx5* regulates genes that show circadian expression in the zebrafish pineal complex. *Nat Genet* 2002;30:117–21. [PubMed: 11753388]
59. Mano H, Kojima D, Fukada Y. Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. *Brain Res Mol Brain Res* 1999;73:110–8. [PubMed: 10581404]
60. Philp AR, Bellingham J, Garcia-Fernandez J, Foster RG. A novel rod-like opsin isolated from the extra-retinal photoreceptors of teleost fish. *FEBS Lett* 2000;468:181–8. Erratum in: *FEBS Lett* 2000 May 4; 473(1):125–6. [PubMed: 10692583]
61. Fitzgibbon J, Hope A, Slobodyanyuk SJ, Bellingham J, Bowmaker JK, Hunt DM. The rhodopsin-encoding gene of bony fish lacks introns. *Gene* 1995;164:273–7. [PubMed: 7590342]
62. Bellingham J, Tarttelin EE, Foster RG, Wells DJ. Structure and evolution of the teleost extraretinal rod-like opsin (*erro*) and ocular rod opsin (*rho*) genes: is teleost *rho* a retrogene? *J Exp Zool B Mol Dev Evol* 2003;297:1–10.
63. Imanishi Y, Yang L, Sokal I, Filipek S, Palczewski K, Baehr W. Diversity of guanylate cyclase-activating proteins (GCAPs) in teleost fish: characterization of three novel GCAPs (GCAP4, GCAP5, GCAP7) from zebrafish (*Danio rerio*) and prediction of eight GCAPs (GCAP1–8) in pufferfish (*Fugu rubripes*). *J Mol Evol* 2004;59:204–17. [PubMed: 15486694]
64. Takechi M, Kawamura S. Temporal and spatial changes in the expression pattern of multiple red and green subtype opsin genes during zebrafish development. *J Exp Biol* 2005;208:1337–45. [PubMed: 15781894]
65. Bridges CD, Liou GI, Alvarez RA, Landers RA, Landry AM Jr, Fong SL. Distribution of interstitial retinol-binding protein (IRBP) in the vertebrates. *J Exp Zool* 1986;239:335–46. [PubMed: 3093630]
66. Si JS, Borst DE, Redmond TM, Nickerson JM. Cloning of cDNAs encoding human interphotoreceptor retinoid-binding protein (IRBP) and comparison with bovine IRBP sequences. *Gene* 1989;80:99–108. [PubMed: 2792773]

67. Liou GI, Ma DP, Yang YW, Geng L, Zhu C, Baehr W. Human interstitial retinoid-binding protein. Gene structure and primary structure. *J Biol Chem* 1989;264:8200–6. [PubMed: 2542268]
68. Wagenhorst BB, Rajendran RR, Van Niel EE, Hessler RB, Bukelman A, Gonzalez-Fernandez F. Goldfish cones secrete a two-repeat interphotoreceptor retinoid-binding protein. *J Mol Evol* 1995;41:646–56. [PubMed: 7490779]
69. Rozen, Steve; Skaletsky, Helen J. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S.; Misener, S., editors. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press; Totowa, NJ: 2000. p. 365-386.

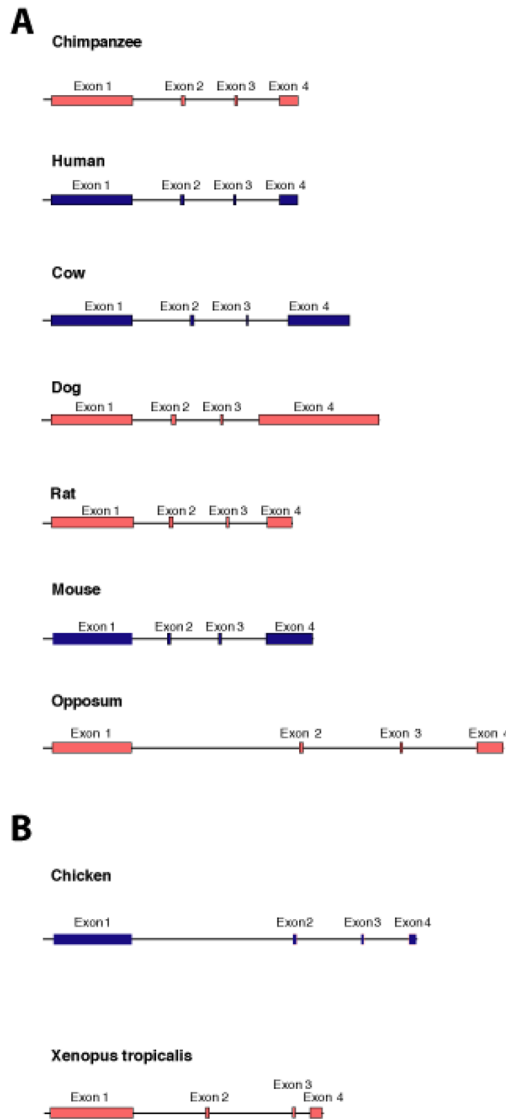


Figure 1.

Tetrapod IRBP gene structures. There is a common gene structure for IRBP among the mammals, birds, and amphibians with a large first exon encoding three full Repeats and parts of the fourth and final Repeat are spread among the first through fourth exons. The blue genes are previously published (human [66,67], mouse [48], cow [26], and chicken [51]), while the red genes are mined from data made available through the courtesy of several large-scale sequencing centers as identified in the Acknowledgments. **A** illustrates the gene structures of the mammals. The dog IRBP gene may have an extra Intron very near the 3' end of the gene, and this is in the 3' UTR. Among the orthologs, the positions of the introns in IRBP are constant, and the coding regions in each exon are the same. **B** shows the IRBP gene structures from non-mammalian tetrapods.

	Donor			Acceptor	
	Exon	Intron		Intron	Exon
Exon 1/Intron A/Exon 2					
chimp	GCCCATGCAG	GTGAGACCCAAGAGA	TAAAACTGGCTGCTCCTCCTGACACTGAGTAGGACCTCCACCTCTTACAG	ATCCCTTCCC	
human	GCCCATGCAG	GTGAGACCCAAGAGA	TAAAACTGGCTGCTCCTCCTGACACTGAGTAGGACCTCCAACCTTTACAG	ATCCCTTCCC	
cow	ACCCATGCAG	GTGAGTACAAGGGA	CTTTCAAATCTTTCCTTTCCTGACATGGAGCTGACCTCCCCTCTTTCAG	ATCCCTTCCC	
dog	GCCCATGCAG	GTGAGACACGCGGTC	CCTGTGCCCGGGCTGAGTGGAGCCCGCAGCCCTGCCCCTGCTCTTGCAG	ATCCCTTCCC	
rat	GCCCATGCAG	GTAAGGTACAAGAGA	CCATACCCACCCCACTCCAGGCACTGCCTGACTCTCCCTCTCTCCAG	ATCCCTTCCC	
mouse	GCCCATGCAG	GTAAGGTACAACAGA	CGCCATCTGTGTGTCTCTGATCTGAGCCAGGCTTCCCCTCTTGCAG	ATCCCTTCCC	
opposum	ATGCAGGTGA	GTCTCCAAAACAGCC	GAAGGAATGCGTTCCATGACTTTTCCAACCTTAATTTGAATTTTGTGCAG	CTTCCCTCTC	
chicken	GCCAAAACAG	GTACAGTTTTTTTTT	AGCTATCAGTGCCAGCAAATGATTTGTTTTGTTTTTCCAACCTCTTTCAG	ATCCCTTCCC	
Xen. Trop	GCCAATGCAG	GTATGTAAGTTTTTA	TTTCTACCAGTTTTTACATGTGCTAAAAATCTGCTCTCAATCTTTTAG	ATCCCATCAC	
Exon 2/Intron B/Exon 3					
chimp	TCGACATGAG	GTCAGTGGCCAGGGG	GCCTCACTGTGAGCTCAGCCCCTGAACAGGCTCTGCTTCCCCTCTTTCAG	GTTCAACATC	
human	TCGACATGAG	GTCAGTGGCCAGGGG	GCCTCACTGTGAGCTCAGCCCCTGAACAGGCTCTGCTTCCCCTCTTTCAG	GTTCAACATC	
cow	TCGACATGAG	GTCAGTGGCCCTGAG	GCTTCCCTGTGGGTTCTGCCCCTTAGTGGGCAACTTTTCCCTTCCCTTCAG	GTTCAACATC	
dog	TCGACATGAG	GTCAGAGTCAGGGGG	GGCCGCCCCATCAGCCCCTGCCCCAGCGGTCCCCTCTCTCCCCGCGAG	GTTCAACATC	
rat	TCGACATGAG	GTCAGTGGGCTAGAC	TTGAGTGCTCTCAGACCAACCCTCAGAGTCATTGTCTGTGCTCCCTTCAG	ATTTAACATC	
mouse	TCGACATGAG	GTCAGTGGGCTGAC	CTCCCTGTGAGCTCAGACGCCCTCCGAGTCATTGGCTGTGCTCCCTTCAG	GTTCAACCTT	
opposum	TTGATATGAG	GTAACATGGGAGGG	TCAGGAGCTTTAATTTGGTTTCTAATGTCTACTAGTTCATGCCAACAG	GTTCAACATA	
chicken	TAGACATGAG	GTAACCTTTGAATG	AGCTATCAGTGCCAGCAAATGATTTGTTTTGTTTTTCCAACCTCTTTCAG	ATCCCTTCCC	
Xen. Trop	TTGATATGAG	GTAGGTTGTCATTTC	TAATATAAAAATAAGACTTATATTCAACAACATTTATTTGTTTACATTTTCAG	ATTTAACATC	
Exon 3/Intron C/Exon 4					
chimp	CAGGTTGTAG	GTACGTGGAGAAGCT	CTCCAGGGTCTGACATGACCCCATCCTGAAGGGCCTTCTGTCTTCCAG	GTGAACGCTA	
human	CAGGTTGTAG	GTACGTGGAGAAGCT	CTCCAGGGTCTGACATGACCCCATCCTGAAGGGCCTTATGTCTTCCAG	GTGAACGCTA	
cow	CAGCTTGAAG	GTGCGTGCAGAAGGC	CCAGCGTGGGATTTGACTCCCTATTTTGAAGGGTCTTGTATCTTGCAG	GTGAACGCTA	
dog	GACTCTGTCA	GTGAGCTCTGGACCC	CAGGGGCCCGGGGACTGCCAGCCGGCTTGAAGGGCCCTGTGTCTTGCAG	GTGAGCGCTA	
rat	CAGCTCACAG	GTACATGAAGCCTTC	GGTGGGAGGATTTCAATTTGGCCCATGCTGAATGGCCCTCTGTCTCCAG	GTGAACGATA	
mouse	CAGCTCACAG	GTACCTGAAGCCTTC	AGGTGGCGGGGCTGATTTGACCCCGTCTGAATGGCCCTCTGTCTCCAG	GTGAACGATA	
opposum	ACAGTCTCAG	GTTGCAAGGTGAGAGG	CAATTACAGTCTCCAGAATCTTATATTTGTCAATTTTGTCTCTTGCAG	GTGAACGATA	
chicken	TAGACATGAG	GTAACCTTTGAATG	AATCTCTGAAAATGGTAAAATTTTGTGATGCTTTTTTGTCTTTACGGCAG	GGGAGAGGTA	
Xen. Trop	CATCTTGTAG	GTATGTAAGTGCACA	TATTCATTTAAATTTATACCGCTGGATGTGCTTATTTTTTCCCTACACAG	GAAATGCTTT	
	Consensus CAR	GTRAGT...	... YYYYYYYYYYNYAG	G	

Figure 2. Alignment and comparison of Donor and Acceptor sites from the tetrapods. This is a comparison to the accepted consensus donor and acceptor sites from human genes. All the introns contain invariant GT and AG dinucleotides at the beginning and end of the intron. Most nearby nucleotides closely match the consensus donor or acceptor motif. The lariat consensus sequence is detected in about half the introns within 50 nucleotides of the acceptor site, and all have several A's, which may function in the absence of a closely matching lariat sequence. These alignments indicate that the splice sites are all strongly conserved in position and sequence, and they all appear to function well.

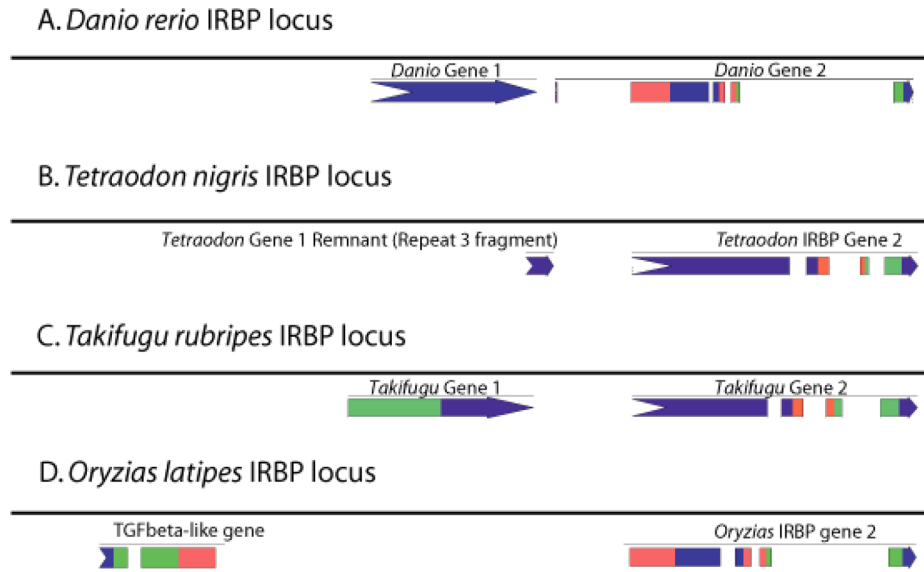


Figure 3. Teleost fish IRBP gene loci. **A:** Zebrafish has two IRBP genes. **B:** *Tetraodon nigroviridis* has a remnant of the upstream gene (Gene 1) and a full copy of Gene 2. **C:** Fugu has two IRBP genes. Panel D: Medaka has a single gene corresponding to Gene 2. In zebrafish and fugu, the two IRBP genes are oriented head to tail, and, across species, the intergenic spacer is about the same size. No intervening genes are found between the two genes in the single locus. The colors originate with the web site computer program called AUGUSTUS [46]. Each exon is marked with two different colors, such that the end of one exon and the beginning of the next exon have the same color. Colors are rotated so that adjacent exons are colored differently (only three colors are needed) and the order of use is blue-red-green, which is then repeated. The first exon of a gene is depicted with an arrow tail at the 5' end, and the last exon is labeled with an arrowhead at the 3' end. For genes that have only a single exon, they are solid blue, with an arrow head and tail. These models are predicted gene structures based on several algorithms. Also, some of the features are small, so that a tiny first exon might not appear to have an arrow tail at the 5' end.

	Donor		Acceptor	
	Exon	Intron	Intron	Exon
Exon 1/Intron A/Exon 2				
zebrafish	TCCTCCAATG	GTAAGATTCTTTTC	AGAACTTTAGGTTTTGAGACAAGCTTACGAATGTTCCTTAACTCCAG	AATCCCACAC
tetraodon	GCCCCCATG	GTGAGACCCATGAAT	TATTTAGCCATTAATAGACATAACAAATGTGTATATCTTCTCCAACAG	GATTATACCC
fugu	GAGGCCCA	GTGCTGCCCCCTATG	ATGGTATGCTTCCAGTCTACTGAAACAAAATGTGTGTTCTTCTCCAACAG	GATTATAGCC
medaka	CCCACCAATG	GTACAACCCACTACA	AAAACCTGTGTAATGGTTGAAATGCTGCTTTGGTTCGCGTCTCTGCAG	GAATACACCC
Exon 2/Intron B/Exon 3				
zebrafish	TTGATCTGAG	GTGATTATTCAATTA	TGGCTATAATTTGATCTGACAACCTGGTATTTTATTTTCGTTGATTCTAG	AAATAACATC
tetraodon	TCGACCTCAG	GTGGGTAATCCCGCT	AAAAATATAGTAAATGAATATATCCTTTTTTGAATGATTTATTTGTCAG	GAACAATGTT
fugu	TTGACCTCAG	GTGGGTAATCCTGCT	AAACAAGATCTGTCTACAGCAAACCTTTTGAGGGTTTTTTTTTATTCGAG	GAACAATGTT
medaka	TCGACCTCAG	GTGGGATGACACCCT	GATTGGGTTTCAAACCTTTCTTCAATCTTTCTGCTCCTGAATTTTCAG	GAACAACGTC
Exon 3/Intron C/Exon 4				
zebrafish	CAGCTCACAG	GTACATTATTGCTGT	GATATATTTTCATATTACAAAACCTGATGCATTAATTCACCTTTTCTCTAG	GCAGGAGATA
tetraodon	GAACCTACTG	GTAAAGAGTTTGTG	CAGACGAGTTAAACGACTTTACCGTTCAGCCCATCCTCCGCAGGCGTGAG	GTACGGCTCC
fugu	GAACCTACCG	GTAAATAAAGATTGT	TGTGCGCTTTGCGTTCCTCATTTAACGCTAACCCATCATGCCCCACAG	GCGTGAGGTA
medaka	GAGCTCACAG	GTGGGAAGGAGAAAA	TATCGAGCTTGAGGCTCTTTACACCGTCAAACATCTCTTCTCCTCTGCAG	GTGAGAGGTA
	Consensus sequence patterns	...CAR	GTRAGT...	...YYYYYYYYYNYAG G

Figure 4.

Alignment and comparison of Donor and Acceptor sites from the teleost fish Gene 2. This is a comparison to the accepted consensus donor and acceptor sites from human genes. At the beginning and end of each intron, respectively, invariant GT and AG dinucleotides are located. Most nearby nucleotides match the consensus donor or acceptor motif. These alignments indicate that the splice sites are all strongly conserved in position and sequence, and they all appear to function correctly.

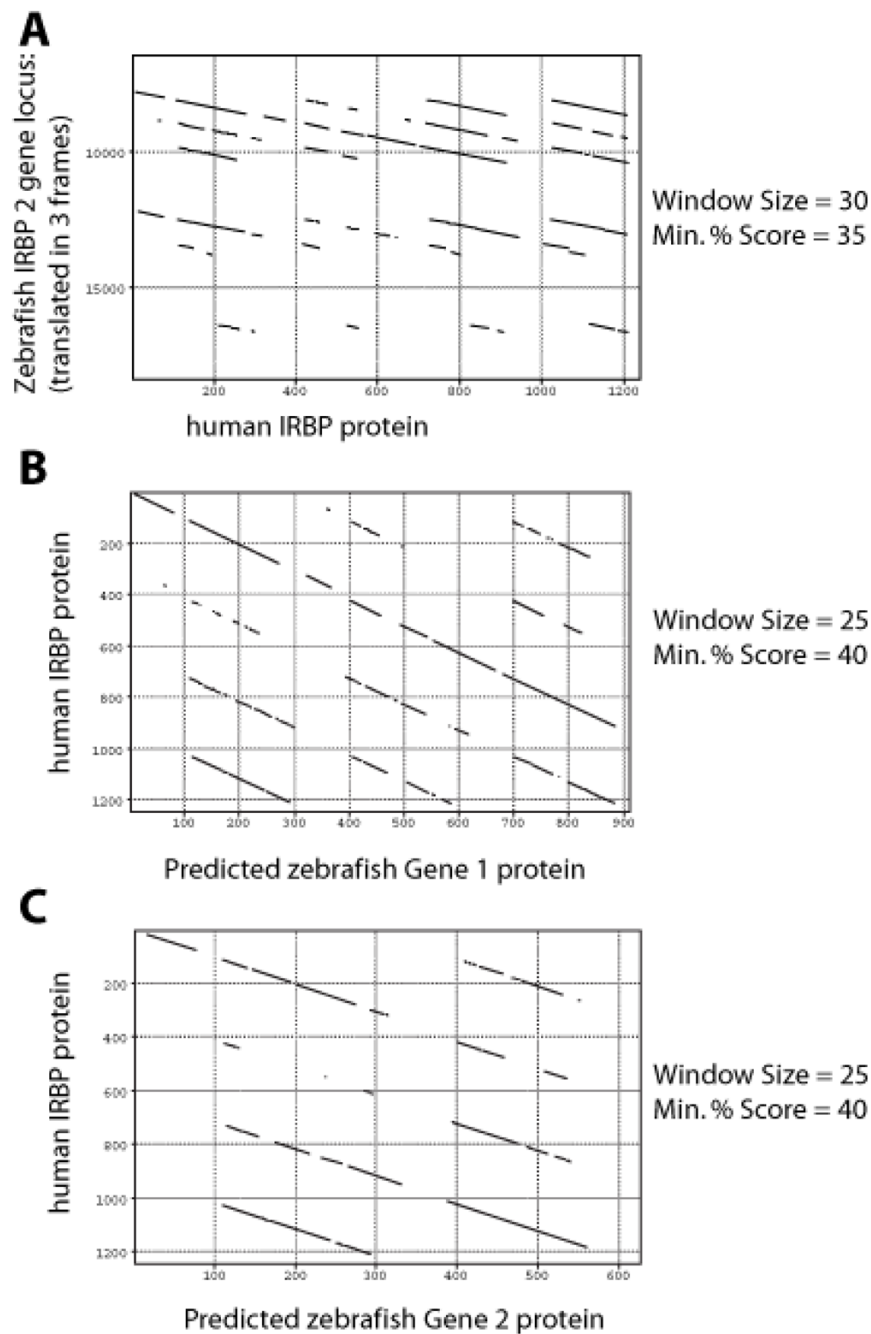


Figure 5.

Dot matrix comparisons showing the existence of a two-gene IRBP locus in zebrafish. **A:** Dot matrix of human protein against the entire zebrafish gene locus. **B:** Human protein against translated Gene 1. **C:** Human protein against Gene 2 protein. The Pustell dot matrix program in MacVector was used employing the pam250 weight matrix and a hash value of 2 in all three panels. Two genes exist in tandem. The genes are oriented head to tail, the genes are close together, the genes are different, with Gene 1 a single exon and three-Repeats long and Gene 2 having four exons, but just two repeats. The optimum alignments to the human IRBP amino acid sequence showed that Gene 1 encodes Repeats 1, 2, and 3, while Gene 2 encoded Repeats 1 and 4.

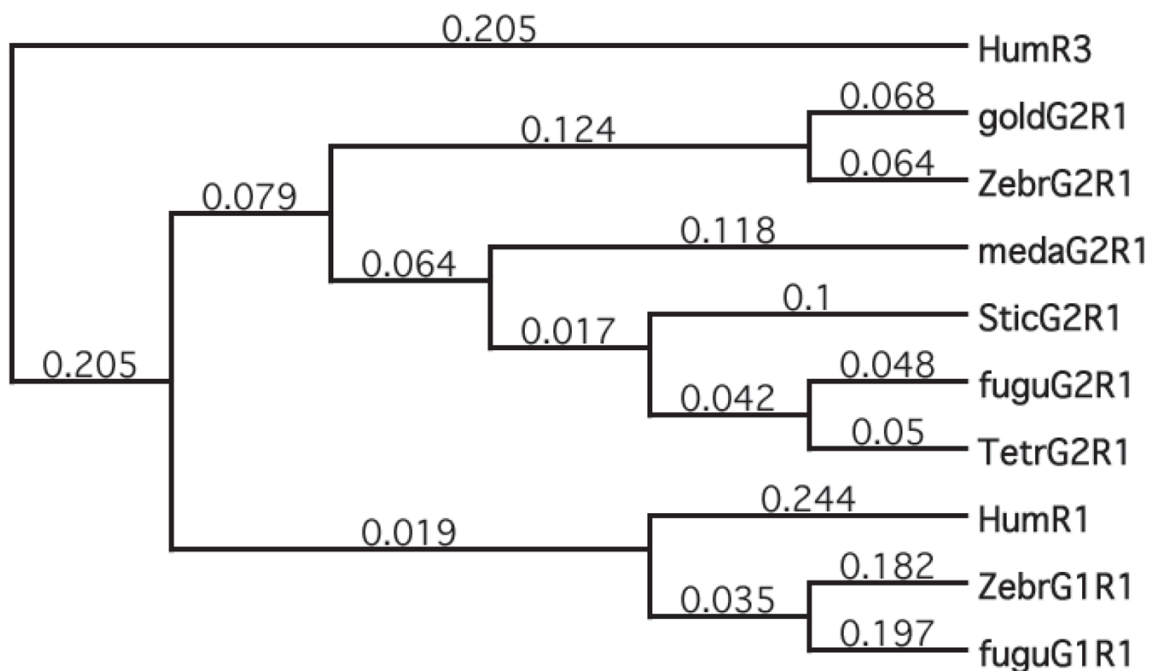


Figure 6.

Phylogenetic analysis of Repeat 1 from IRBP. Evolutionary relationships among fish and human Repeat 1 orthologs and paralogs are illustrated. MacVector version 9.0 was used to build a cladogram of the Repeat 1-like amino acid sequences of the teleost fish and the human Repeat 1. Human Repeat 3 was used as an outgroup for the purpose of rooting this tree. The illustrated tree was built using the neighbor-joining method. Whether using the UPGMA or neighbor-joining method, with uncorrected or Poisson-corrected distances, random or systematic tie-breaking for gaps, in MacVector version 9.0, all the phylogenetic reconstructions gave the same overall structure. A consensus tree from 1000 replications of bootstrapping was calculated, and it was identical to the structure of the best tree, which is illustrated in this image. Among 1000 replications of resampling, all seven relevant nodes occurred frequently, with four appearing in 100% of trees, two nodes occurring in 96% of trees, and one node occurring in 79% of trees. All the nodes are resolved in the illustrated tree. The results suggest a paralogous relationship between the Repeat 1 sequences from Genes 1 and 2 in zebrafish and fugu. The closer relationship of human Repeat 1 to the fugu and zebrafish Gene 1 orthologs than to Gene 2 and the illustrated divergence at the leftmost internal node of the paralogs on this cladogram suggests that the paralogous Genes 1 and 2 arose early or at the origin of the teleosts. This tree is consistent with the already known divergence radiation of the teleost fish about 350 Mya.

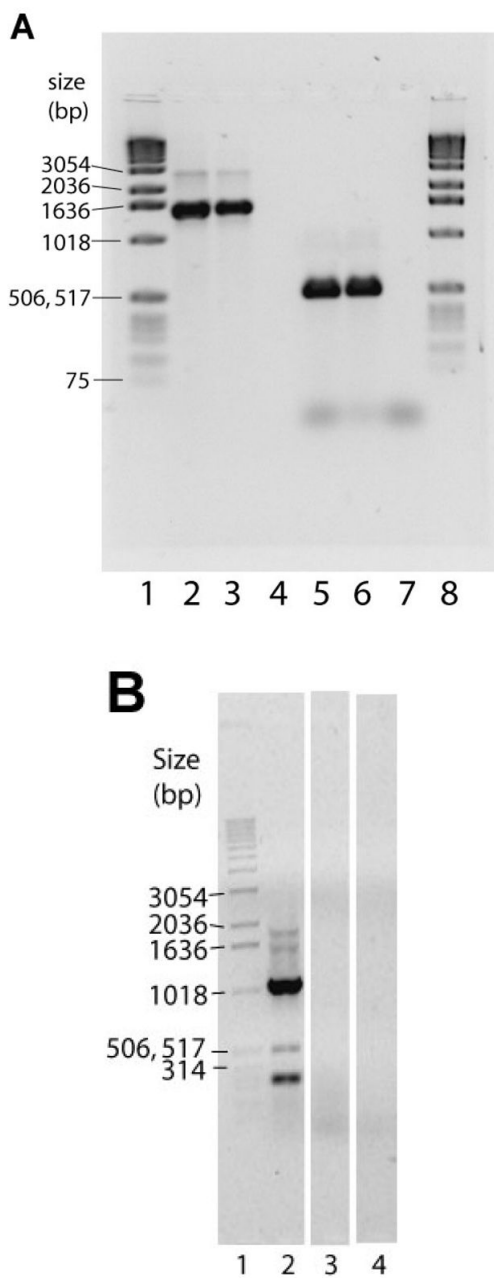


Figure 8.

RT-PCR from Gene 1 or Gene 2-specific primers. **A:** Expression of Gene 1 and Gene 2 in zebrafish tissues. Amplicon size was measured on a 1.0% agarose gel (and run in 1X TAE buffer) stained with 1X SYBR Green. Lanes 1 and 8, 0.5 μ g 1 kb DNA ladder (Invitrogen, Carlsbad, CA). Lane 2, 96 h zebrafish whole larval RNA primed with gn1fwd and gn1rev (Table 2). Lane 3, adult zebrafish eye RNA primed with gn1fwd and gn1rev primers. Lane 4, no RNA control, primed with gn1fwd and gn1rev. Lane 5, 96 h whole larval RNA with gn2 primers, F3 and R4 (Table 2). Lane 6, adult zebrafish eye RNA, primed with gn2 primers, F3 and R4. In the absence of RNA, the F3 and R4 primers amplified no products (data not shown). Lanes 2 and 3 had a product of about 1500 bp, and the size expected from the Gene 1 DNA sequence was 1529 bp. Lanes 5 and 6 had a product of about 530 bp and an expected size based

on the sequence of 532 bp. These results indicate that both genes are transcriptionally active in both zebrafish adult eye and larval whole bodies. Panel B: Expression of an RNA transcript spanning Gene 1 and Gene 2. Lane 1, 1 kb ladder (Invitrogen). Lane 2, an RT-PCR product spanning across the intergenic region using primers F1 and R3 (Table 1) was found with a size estimated to be 1000 bp (96 h larval RNA). Lane 3, same as Lane 2 in the absence of reverse transcriptase. Lane 4, no RNA control.

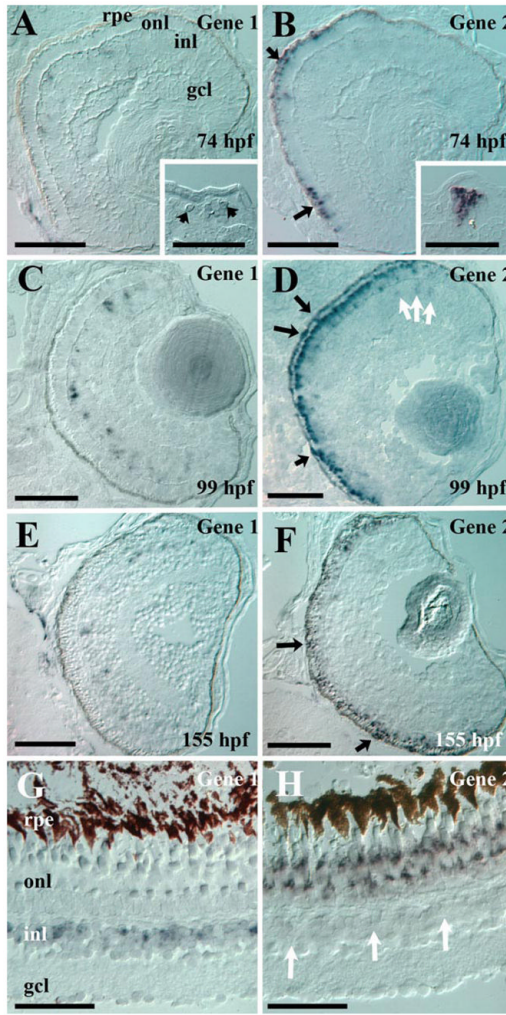


Figure 9. In situ hybridizations. Zebrafish IRBP Gene 1 and Gene 2 are differentially expressed. Retinal cryosections were obtained from 74 hpf (A, B) *albb4* zebrafish embryos, 99 hpf (C, D) and 155 hpf (E, F) *albb4* zebrafish larvae, and adult Tue zebrafish eyes (G, H) and were hybridized with Gene 1- (A, C, E, G) and Gene 2- (B, D, F, H) specific probes. Note that *albb4* zebrafish do have low levels of pigment in the RPE; dark arrows show Gene 2 hybridization. White arrows show weak expression of Gene 2 sporadically in the INL. Insets in A and B shows Gene 1 and Gene 2 expression in developing pineal; contrast was enhanced to show weak expression of Gene 1. Scale bars represent 50 μ m; rpe represent retinal pigmented epithelium, onl represent outer nuclear layer; inl represent inner nuclear layer; gcl represent ganglion cell layer.

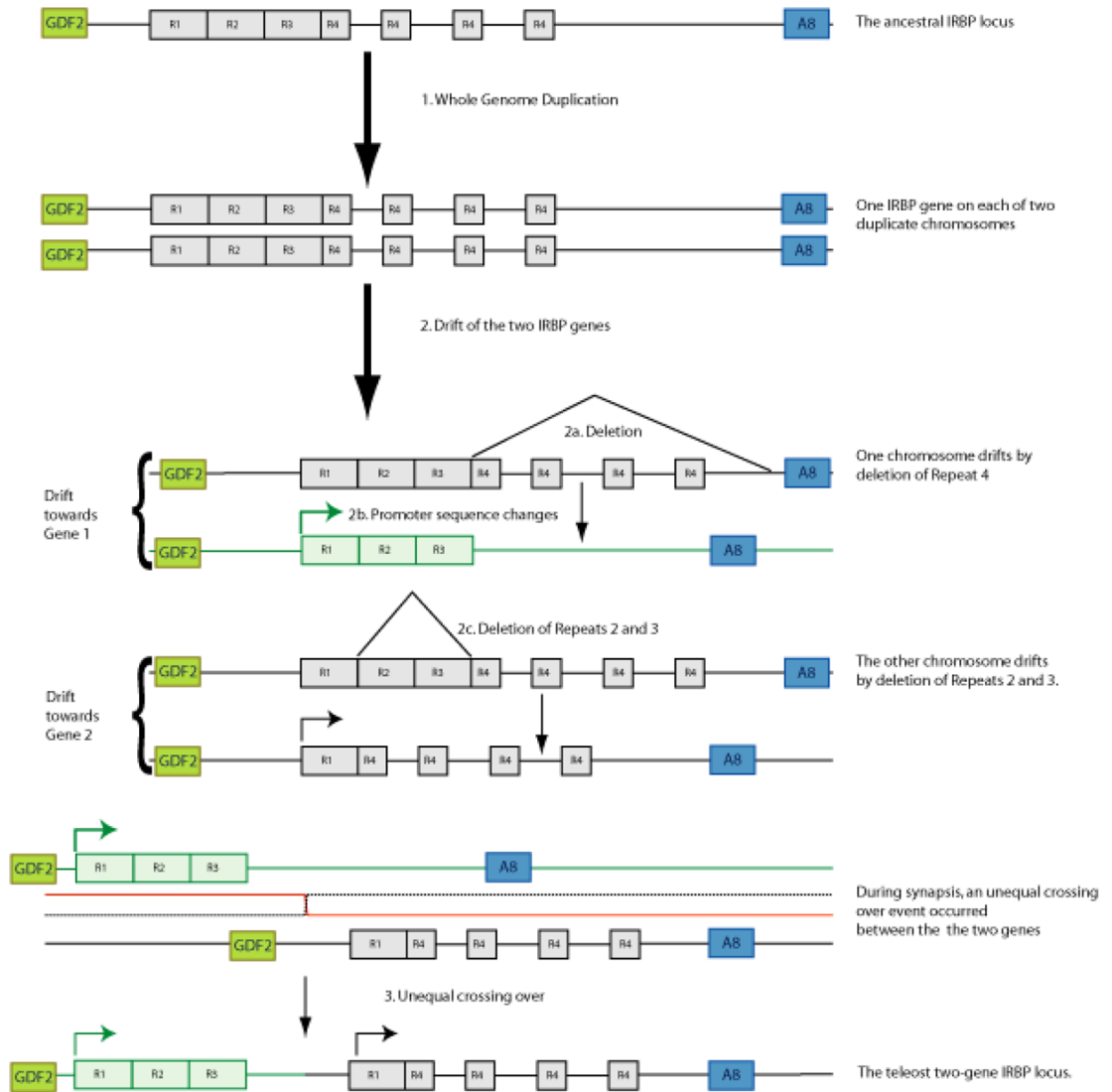


Figure 10. A model of the origin of two IRBP genes in the teleosts. The presumptive ancestral gene in fish, predating the appearance of the teleosts is a gene much like the present day tetrapod gene with four Repeats encoded in a four-exon gene, with the first three repeats and part of the fourth found in the first exon. When the teleosts arose in Step 1, the whole genome was duplicated. In Step 2, the two IRBP genes drifted, deleting different parts of the gene. In the gene destined to become Gene 1, the 3' end of the gene was lost deleting the three introns, exons 2–4 and part of exon 1 (Step 2a), and the promoter underwent some alterations (Step 2b), so that it would be expressed in some ganglion cells and selected cells in the inner nuclear layer. The gene destined to become Gene 2 underwent an internal deletion, losing the parts of Exon 1 that encoded Repeats 2 and 3 (Step 2c) but Gene 2 did not undergo any significant changes in its promoter, retaining expression in photoreceptor cells and RPE cells. In Step 3, unequal crossing over between the two chromosomes would form the current teleost two-gene locus. Synteny is preserved in this model. GDF2 represents the growth differentiation factor 2 gene, and A8 represents the Annexin 8 gene.

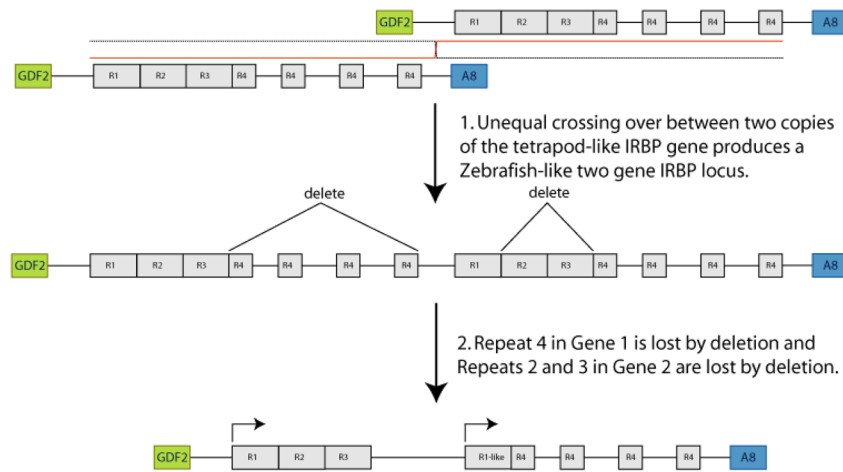


Figure 11.

A model of teleost IRBP locus evolution based on external unequal crossing over. An unequal crossing over event between the two chromatids of the IRBP gene leads to one chromatid bearing two copies of the IRBP gene. In subsequent steps, Repeat 4 is lost from the first gene and Repeats 2 and 3 are lost from the second gene. These steps produce a two-gene locus in head-to-tail orientation. Synteny is preserved in this model. GDF2 represents the growth differentiation factor 2 gene, and A8 represents the Annexin 8 gene.

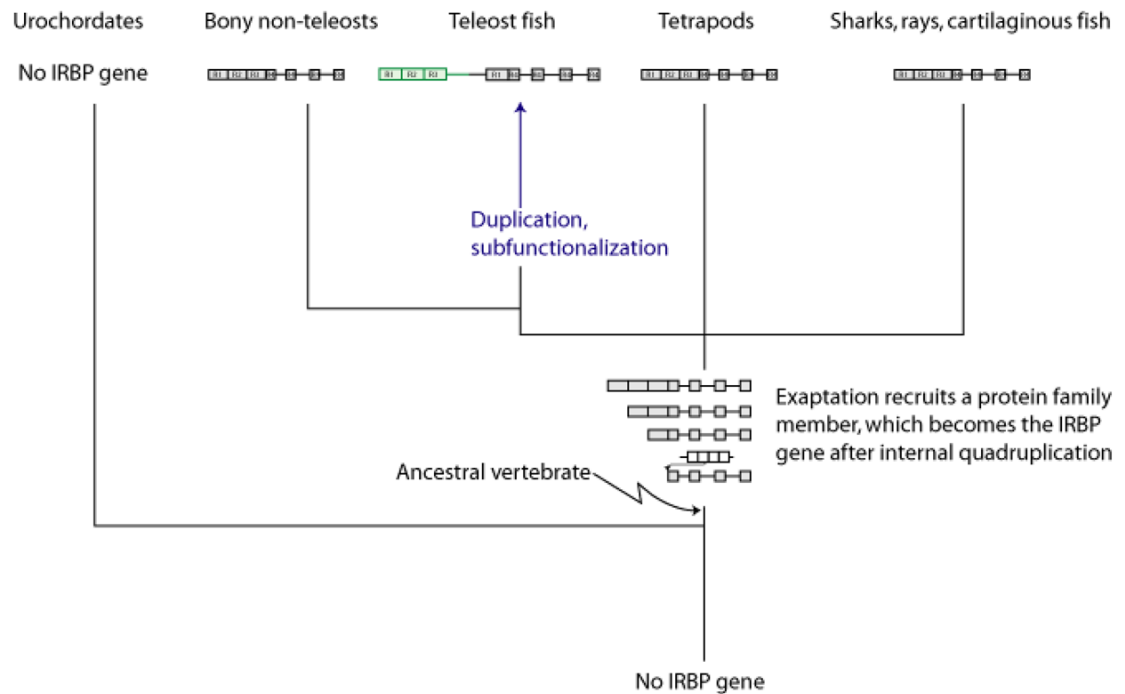


Figure 12. Current model of the origins and evolution of the IRBP gene. The IRBP gene probably arose coincident with or shortly after the vertebrates diverged from the urochordates. This is the time period when the neural crest and the skull originated. No major changes are proposed in the steps resulting in the early internal quadruplication of the IRBP gene, but the time frame is now bounded, as the IRBP gene appears to be absent from the urochordates. This model is based on Borst et al. [26] and Rajendran et al. [27]. The major change is the addition of a two-gene IRBP locus in the teleosts.

Table 1

GenBank Accession or scaffold numbers of the sequences used in this study

Species	GenBank Accession or scaffold number	Submitter or reference
Tetrapods:		
Chimpanzee	NW_001222167	Washington University Genome Sequencing Center, Pan_troglodytes-2.1, March 2006
human	NM_002900	[66,67]
cow	M20748	[26]
Dog	NW_876311	The Broad Institute of MIT/Harvard and Agencourt Bioscience, 10 May
rat	NW_047469	chromosome 16 genomic contig, version 3.4 'November 2004 Update' Baylor College of Medicine Human Genome Sequencing, part of the Rat Genome Sequencing Consortium
mouse	AY682090	[48]
opposum	AAFR03024396	03-JAN-2006, The Broad Institute of MIT and Harvard.
monkey (macacca mulatta)	NW_001124202	Mmul_051212, February 2006 Macaca mulatta Genome Sequencing Consortium
chicken	NW_060392	[51]
Xenopus tropicalis	scaffold 653	DOE Joint Genome Institute, v4.1
Teleost Fish:		
zebrafish	CR450691, X85957	Wellcome Trust Sanger Institute [27]
Takifugu rubripes	scaffold 3	Version 4, June 2005, The International Fugu Genome Consortium
Tetraodon nigroviridis	SCAFFOLD 10211	Chromosome 2, The Tetraodon nigroviridis V7 assembly (February 2004) from Genoscope and the Broad Institute of MIT and Harvard
medaka (Oryzias latipes)	scaffold 1465	National Institute of Genetics (NIG) and the University of Tokyo (October 2005)
Goldfish (Carasius auratus)	X80802	[68]
Gasterosteus aculeatus (three spined stickleback)	DN689753	The Stanford Human Genome Center (submitted 2003), unpublished ESTs.
	DN683402	
	DN695303	
	DN686722	
	DN680257	
	DN684040	
	DN694734	
	DN682976	
	DN691753	
	DN685347	
	DN681402	
	DN686085	
	DN689752	
	DN680256	
	DN685346	
	DN686721	

<u>Species</u>	<u>GenBank Accession or scaffold number</u>	<u>Submitter or reference</u>
	DN686084	
	DN683401	
	DN691752	
	DN684039	
	DN691083	
	DN681401	
	DN694733	
	DN695302	
	DN691082	
Pimephales promelas (fathead minnow)	DT130195	DOE Joint Genome Institute, unpublished (12-JAN-2006)
	DT138893	
	DT138894	
	DT140312	
	DT180381	
	DT182368	
	DT191533	
	DT191534	
	DT191534	

All the sequences of the tetrapods are readily obtained by Blast searches of the public databases. Four of the fish (fugu, zebrafish, medaka, and Tetraodon) are similarly readily obtained from the latest versions of the whole genome sequences. The goldfish sequence was obtained from GenBank. We assembled a sequence from stickleback and fathead minnow sequences of several EST sequences, which are identified by accession number. We emphasize that we did not conduct the sequencing of stickleback or the fathead minnow ourselves.

Table 2

Primer sets for RT-PCR and for the construction of in situ probes

A. Primers for RT-PCR for Gene 1-specific mRNA in situ probe			
<u>Primer ID</u>	<u>Sequence (5'-3')</u>	<u>T_m</u>	<u>Expected product size (bp)</u>
gn1fwd	AGGAGGATTTGGCGGAAAA	65.02	1529
gn1rev	AACTGCCTGATTGGGAATCGAA	65.05	
B. Primers for RT-PCR for Gene 2-specific mRNA in situ probe			
<u>Primer ID</u>	<u>Sequence (5'-3')</u>	<u>T_m</u>	<u>Expected product size (bp)</u>
F3	gatgttcacgcgctcatca	63.32	532
R4	gcagtgctgaatggctgat	58.83	
C. Primers for RT-PCR spanning the intergenic region between Gene 1 and Gene 2			
<u>Primer ID</u>	<u>Sequence (5'-3')</u>	<u>T_m</u>	<u>Expected product size (bp)</u>
F1	ttacgactcttcagatgcttg	55.2	1098
R3	ggtttccaactgagctgcttc	56.8	

These primers were used the reverse transcriptase-coupled PCR experiments. Other primers were used in the construction of probes for in situ hybridization. Primer3 [69] was used to predict suitable primers. T_m represents the melting temperature.

Table 3

Percent identities and scores comparing Repeat 1 protein orthologs and paralogs

		Repeat 1 gene and species					
		Z1	Z2	F1	F2	HR1	HR2
A. Orthologs and paralogs compared in zebrafish and fugu							
Repeat 1 gene and species	Z1	Z2	F1	F2	HR1	HR2	
zfish Gene 1 Repeat 1 (Z1)	---	337	392	343	326		
zfish Gene 2 Repeat 1 (Z2)	52%	---	302	416	289		
fugu Gene 1 Repeat 1 (F1)	59%	48%	---	306	312 (score bits)		
fugu Gene 2 Repeat 1 (F2)	52%	63%	49%	---	297		
human Repeat 1 (HR1)	53%	42% percent identity	51%	47%	---		
B. Repeat 1 orthologs compared among teleost fish							
Repeat 1 gene and species	Repeat 1 gene and species						HR1
	Z2R1	F2R1	MR1	TR1	GRI	SRI	
zfish Gene 2 Repeat 1 (Z2R1)	--	416	384	406	543	402	289
fugu Gene 2 Repeat 1 (Z2R1)	63%	--	488	565	408	515	179
Medaka Repeat 1 (MR1)	59%	74%	--	488	379	491	295 (score bits)
Tetraodon Repeat 1 (TR1)	63%	89%	75%	--	379	508	298
Goldfish Repeat 1 (GRI)	86%	62%	59%	61%	--	400	281
Stickleback Repeat 1 (SRI)	61%	79%	76%	78%	61%	--	295
Human Repeat 1 (HR1)	42%	47% percent identity	45%	47%	42%	46%	--

A: Data suggest that the zebrafish and fugu IRBP loci as a pair of orthologs (Z1-F1 and Z2-F2) diverged more recently than Gene 1 and Gene 2 paralogs (Z1-Z2 and F1-F2). The paralogs Z1-F2 and Z2-F1 are consistent with the paralogs Z1-Z2 and F1-F2, in percent identity and score, suggesting that there has been no recent gene conversion. The greater conservation of the orthologs than the paralogs suggests that the divergence of Gene 1 from Gene 2 was an early event in teleost evolution, as the divergence of fugu and zebrafish was in the moderate to distant past in the history of the teleosts. B: Pairwise comparisons between the indicated two versions of Repeat 1 were conducted using BL2SEQ at the Biology Workbench (San Diego Supercomputer Center). The percent identity is indicated in the lower left half of each table. The raw bit score is indicated in the upper right of the table.