Subfunctionalization of Duplicate *mitf* Genes Associated With Differential Degeneration of Alternative Exons in Fish

Joachim Altschmied,*,[1,2] Jacqueline Delfgaauw,*,[1] Brigitta Wilde,* Jutta Duschl,* Laurence Bouneau,[†] Jean-Nicolas Volff*,[3] and Manfred Schartl*

**Department of Physiological Chemistry I, Biocenter (Theodor-Boveri Institute), University of Würzburg, D-97074 Würzburg, Germany and [†]Genoscope, F-91057 Evry, France*

Manuscript received December 13, 2001
Accepted for publication February 22, 2002

ABSTRACT

The microphthalmia-associated transcription factor (MITF) exists in at least four isoforms. These are generated in higher vertebrates using alternative 5′ exons and promoters from a single gene. Two separate genes (*mitf*-m and *mitf*-b), however, are present in different teleost fish species including the poeciliid Xiphophorus, the pufferfishes *Fugu rubripes* and *Tetraodon nigroviridis,* and the zebrafish *Danio rerio*. Fish proteins MITF-m and MITF-b correspond at both the structural and the expression levels to one particular bird/mammalian MITF isoform. In the teleost lineage subfunctionalization of *mitf* genes after duplication at least 100 million years ago is associated with the degeneration of alternative exons and, probably, regulatory elements and promoters. For example, a remnant of the first exon specific for MITF-m is detected within the pufferfish gene encoding MITF-b. Retracing the evolutionary history of *mitf* genes in vertebrates uncovered the differential recruitment of new introns specific for either the teleost or the bird/mammalian lineage.

GENE duplications are important events in evolution (Ohno 1970). One major advantage is that one of the versions of the ancestral genes is free to diverge independently of the selective forces that maintain the original function. This may lead to one of three scenarios. First, and obviously most frequently, one copy may be inactivated by accumulation of random mutations and become nonfunctional ("defunctionalization"). Second, one of the genes may diverge to a state where it acquires a new function ("neofunctionalization"). Third, both genes may be altered in a way such that the combined activity of the two genes together fulfills the task of the ancestral gene in a complementary fashion ("subfunctionalization"; Force *et al.* 1999).

Fish are a group of organisms especially suited for the study of such events, because it became evident that they have more duplicated genes than do the other groups of vertebrates (Wittbrodt *et al.* 1998; Meyer and Schartl 1999). Within the three consequences of gene duplication, subfunctionalization has attracted a lot of interest and paradigmatic cases have been described of genes where two genes in fish perform together the function of their common tetrapod ortholog (Force *et al.* 1999).

Two different isoforms of the microphthalmia-associated transcription factor (MITF) that appear to fulfill the criteria of subfunctionalization (Lister *et al.* 2001) were recently described in zebrafish. MITF is a member of the basic helix-loop-helix leucine zipper (bHLH-Zip) protein family (Hodgkinson *et al.* 1993; Hughes *et al.* 1993). It plays a central role in the differentiation and/or survival of melanocytes, which is reflected in the correlation between a number of mutant pigmentation phenotypes and specific mutations in the *mitf* gene (Steingrimsson *et al.* 1994; Lister *et al.* 1999; Hallsson *et al.* 2000). Different mutations in the human gene can lead to Waardenburg syndrome type 2 (Tassabehji *et al.* 1994) or Tietz syndrome (Smith *et al.* 2000), both of which are associated with pigmentary changes.

In mammals and birds, several isoforms of MITF that are encoded by a single gene are known. The different proteins are generated by the use of alternative promoters and 5′ exons (Yasumoto *et al.* 1998; Udono *et al.* 2000). The zebrafish data, on the other hand, indicate the presence of different genes for at least two isoforms in this organism (Lister *et al.* 1999, 2001).

The mammalian and avian MITF-m isoforms have a unique N terminus of 11 amino acids not found in the other variants a, c, and h. These last three isoforms have a common exon, B1b, directly upstream of exon 2, but differ in their N terminus, encoded by the exons a, c, and h, respectively. Differences are not limited to the

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF475090 and AF475091.

[1]These authors contributed equally to this work.

[2]*Present address:* Molecular Hematology, University of Frankfurt Medical School, D-60590 Frankfurt, Germany.

[3]*Corresponding author:* Department of Physiological Chemistry I, Biocenter (Theodor Boveri Institute), University of Würzburg, Am Hubland, D-97074 Würzburg, Germany.
E-mail: volff@biozentrum.uni-wuerzburg.de

primary structure of the protein, but are also found at the transcriptional level. Whereas *mitf*-m is expressed exclusively in melanocytes and melanoma cells, the other isoforms are expressed in a much wider spectrum of cells (Amae *et al.* 1998; Udono *et al.* 2000). Convergent evidences from different organisms including human indicate that MITF-m is the isoform involved in melanocyte differentiation (Vachtenheim and Novotna 1999; Yajima *et al.* 1999). The MITF-a isoform most likely exerts its major function in the retinal pigment epithelium (RPE).

In zebrafish two cDNAs coding for MITF have been isolated, which differ not only in their 5′ terminal exons but over the entire length. These cDNAs originate from two separate genes. One, *mitf*-a is required for the development of neural crest-derived pigment cells, while the other, *mitf*-b, is expressed in the RPE and can rescue MITF-a function in the eye, but not in the neural crest-derived pigment cells. This situation clearly demonstrates the use of two *mitf* genes in zebrafish as an alternative genomic strategy to the subfunctionalization through alternative promoter/5′ exon usage of birds and mammals.

For a full evolutionary understanding of this peculiar situation, in addition to the gene function and cDNA sequence data, information on the structure of the corresponding genes is required, which is not available from fish so far. Another important question that was evoked from the zebrafish data concerns the origin and fate of the two *mitf* genes. It has been proposed that the source for most of the duplicated genes in fish was an ancient whole genome duplication in the lineage leading to modern day teleosts (Amores *et al.* 1998; Postlethwait *et al.* 2000; Taylor *et al.* 2001). This is suggested by the presence of large syntenic segments on different chromosome pairs in zebrafish. Zebrafish *mitf* genes, located respectively in linkage groups 6 and 13, are apparently not included in such large regions of synteny (Lister *et al.* 2001). Although additional genomic rearrangements might have disrupted syntenic regions after genome duplication, this result is at least equally consistent with a more local event of gene duplication. Thus, the duplication and subfunctionalization of *mitf* genes could be an exceptional situation in zebrafish but not a mechanism of general importance in fish.

To perform a broader evolutionary and functional survey of this pair of duplicate genes in the teleost lineage, we have analyzed *mitf* genes in the poeciliid Xiphophorus, which comes from a totally different branch of the phylogenetic tree of teleosts. The genomic structure of *mitf* genes was also determined from the almost completely sequenced genome of *Fugu (Takifugu) rubripes* and from an advanced version (6x coverage) of the genome of the related freshwater pufferfish *Tetraodon nigroviridis*.

We show the presence of two *mitf* genes in all these different teleost species, indicative of a duplication at least 100 million years ago. We find that the *mitf* genes of Xiphophorus are differentially expressed, supporting the interpretation that these genes have experienced an ancient subfunctionalization associated with differential degeneration of specific regulatory sequences and use of alternative exons in each of the duplicate genes.

MATERIALS AND METHODS

Experimental animals and cell lines: All fish used for this study were bred under standard conditions (Kallman 1975). *Xiphophorus maculatus* strain I is homozygous for the golden mutation; strain II represents wild-type *X. maculatus*, origin Rio Jamapa, Mexico, and is derived from the Jp163A strain; and strain III stands for ornamental *X. maculatus*. *X. helleri* is a wild-type strain from Rio Lancetilla, Guatemala.

The embryonic epithelial cell line A2 (Kuhn *et al.* 1979) and the melanoma cell line PSM (Wakamatsu 1981), both from Xiphophorus, were cultured as described previously (Baudler *et al.* 1997).

DNA and RNA experiments: Fish genomic DNA was extracted according to Schartl *et al.* (1995). Standard PCRs were performed with 200 ng genomic DNA as template. For RT-PCR, 1 μg of total RNA isolated with TRIzol (GIBCO BRL, Gaithersburg, MD) was reverse transcribed with Super Script II (GIBCO BRL) in the presence of random hexamers. X*mitf*-m-specific cDNA was amplified in RT-PCR using primers MiEx1 for (5′-TGCTGGAGATGCTGGAATACAG-3′) and MiEx9rev3 (5′-GACGGTGAGACCGTGAGCC-3′); X*mitf*-b-specific cDNA was amplified with MiExBfor3 (5′-CGCTATCAATGTCAGT GTCCC-3′) and MiEx9rev3. To obtain full-length cDNA sequences, the 5′ and 3′ rapid amplification of cDNA ends (RACE) systems from GIBCO BRL were used. PCR fragments were cloned into pUC18 using the SureClone kit (Amersham Pharmacia Biotech) or into pCR II-TOPO (Invitrogen, San Diego). DNA sequencing was done using the ThermoSequenase fluorescent labeled primer cycle sequencing kit with 7-deaza-dGTP and run on an ALFexpress automated laser fluorescent sequencer (Amersham Pharmacia Biotech), or using the CEQD TCS dye terminator cycle sequencing kit and run on a CEQ 2000XL DNA sequencing system (Beckman-Coulter).

Southern blotting was done according to standard protocols. Hybridization was performed in 50% formamide at 42° and the filters were washed with 0.1× SSC/1% SDS at 68°. Screening of the PSM cDNA λ phage library was performed using a PCR-based phage selection procedure (Israel 1993). The cDNA-containing plasmid was excised *in vivo* (Stratagene, Lambda-ZAP Express library).

Sequence analysis: Multiple sequence alignments were generated using "PileUp" of the GCG Wisconsin package (Version 10.0, Genetics Computer Group, Madison, WI). Phylogenies were determined with PAUP* (Swofford 1989) by bootstrap analysis using maximum parsimony (100 replicates) and neighbor joining (Saitou and Nei 1987; 1000 replicates). Maximum-likelihood analysis was performed by quartet puzzling (10,000 puzzling steps; Strimmer and von Haeseler 1996) with PAUP*. The pairwise number of synonymous (K_s) and nonsynonymous (K_a) substitutions per site between two aligned sequences was estimated with "Diverge" of GCG using an unambiguous alignment of DNA sequences [390 nucleotides (nt) in length]. Gene structure was analyzed using programs available on the NIX server (http://menu.hgmp.mrc.ac.uk/menu-bin/Nix).

Isolation of Xiphophorus *mitf* cDNAs: To obtain Xiphophorus sequences coding for MITF, RNA from the melanoma cell

line PSM was reverse transcribed and the resulting cDNA was used as template in a touchdown PCR using degenerate primers MiEx8for (5′-TGMGSTGGAAYAARGGMACC-3′) and MiEx8rev (5′-CTGWAYWCKGAGCARSARRTG-3′). These primers were derived from the region encoding the helix-loop-helix domain, which is highly conserved between the known vertebrate MITF proteins. A product of 140 bp was obtained, which unequivocally codes for a part of MITF different from all other vertebrate MITFs and, thus, most likely codes for Xiphophorus MITF. A 902-bp cDNA fragment encoding parts of a different isoform (XMITF-b) was obtained from the amplification of liver cDNA with the primers MiExBfor3 (5′-CGC TATCAATGTCAGTGTCCC-3′) and MiEx9rev3 (5′-GACGGT GAGACCGTGAGCC-3′).

RESULTS

Duplicate *mitf* genes in Xiphophorus: To isolate *mitf* cDNA(s) from Xiphophorus, nondegenerate primers derived from the highly conserved small fragment within the helix-loop-helix region (see MATERIALS AND METHODS) were used to screen a PSM melanoma cell line cDNA library by PCR. A cDNA encompassing a complete MITF open reading frame was obtained, which encodes the homolog of the mammalian melanocyte-specific isoform MITF-m (XMITF-m) as indicated by the presence of the melanocyte-specific first exon.

Using a degenerate primer derived from the mammalian exon B1b (AMAE *et al.* 1998) in combination with a primer from exon 9 common to all isoforms, we obtained from liver RNA a *mitf*-cDNA distinct from the X*mitf*-m cDNA, which was designated X*mitf*-b in analogy to the corresponding gene from zebrafish. 5′ and 3′ RACE were used to complete the sequence information. Differences between X*mitf*-m and X*mitf*-b were not limited to the 5′ exons, but were found along the whole sequence (Figure 1). The cDNAs showed only 72% identity at the nucleotide level (data not shown), indicating that they originated from separate genes. Southern blot analyses with probes from both cDNAs confirmed this interpretation (Figure 2).

The predicted proteins show all the hallmarks of their counterparts from higher vertebrates (Figure 1). The basic helix-loop-helix region in the central part is identical to the mammalian protein with the exception of a single lysine-to-arginine substitution in a position where a methionine residue is found in the pufferfish MITF-m proteins. The leucine zipper motif directly adjacent serving as a dimerization interface is also well conserved. Mapping studies had revealed the presence of two activation domains in the mammalian protein, one strong acidic activator region in the N-terminal half of the molecule and one threonine-rich region close to the C terminus with modest transactivation properties (TAKEDA *et al.* 2000). While the first region is nearly identical in the fish proteins with all acidic residues present, the second region is poorly conserved.

Expression of the X*mitf* isoforms in tissues from adult Xiphophorus fish and in two established cell lines was analyzed by RT-PCR (Figure 2). X*mitf*-b seems to be expressed ubiquitously, as specific transcripts were present in all cell types analyzed. The actual concentrations are fairly variable with highest levels found in muscle. X*mitf*-m RNA, however, was detected only in tissues containing melanocytes, like eyes and skin, and in melanomas as well as in the melanoma cell line PSM, thus paralleling the expression pattern of *mitf*-m in mammals. In addition, the levels of X*mitf*-m RNA in the melanoma-derived samples are considerably higher than levels of X*mitf*-b transcripts.

Duplicate *mitf* genes in pufferfish genomes: Two different *mitf* genes could be identified in sequences generated by the Japanese pufferfish *F. rubripes* genome project (ELGAR *et al.* 1996, 1999; http://www.jgi.doe.gov/programs/fugu/fugu_mainpage.html, http://fugu.hgmp.mrc.ac.uk/blast/): *mitf*-b (accession FT:T000263 scaffold_263, length 82968 nt), and *mitf*-m (accession FT:T008421 scaffold_8422 length 10919 nt). Both *mitf* genes could be reconstructed from the freshwater pufferfish *T. nigroviridis* as well (http://www.genoscope.cns.fr, CROLLIUS *et al.* 2000; FISCHER *et al.* 2000; sequences available on request).

In both pufferfishes, the nucleotide identity between exonic sequences from the two *mitf* genes was again only ~72%. All four genes contain introns at similar positions, indicating that the duplication of the *mitf* gene occurred at the genomic level rather than through retrotransposition of *mitf* mRNA (Figure 3). The prediction of intron placement and exon sizes in the pufferfishes match the gene structure of *mitf*-m in Xiphophorus, which was determined from the size and/or sequence of genomic PCR products (data not shown).

Phylogeny and evolution of MITF sequences: In contrast to the situation observed in higher vertebrates, phylogenetic analyses supported the presence of two distinct groups of orthologous MITF sequences in teleost species having diverged at least 100–120 million years ago: the MITF-b group, including proteins homologous to the MITF-b proteins from the zebrafish *Danio rerio* (LISTER *et al.* 2001) that contain sequences encoded by exons 1a and 1b (structurally related to the MITF-a isoform from higher vertebrates), and the MITF-m group, including proteins homologous to the zebrafish MITF-m (LISTER *et al.* 1999, 2001) that contain exon 1m and thus are structurally more related to the MITF-m isoform from mammals and birds (Figures 1 and 4). Both groups contain *bona fide* MITF proteins that do not correspond to fish versions of the related TFE3 transcription factor (REHLI *et al.* 1999). Maximum parsimony (bootstrap value 86%) but neither neighbor joining nor maximum-likelihood analyses supported the existence of a fish-specific group containing both MITF-b and MITF-m proteins.

Since duplicated gene copies without evolutionary constraints can frequently evolve as pseudogenes, we compared the number of synonymous (K_s) *vs.* nonsynonymous (K_a) substitutions in the *mitf* genes. In all com-

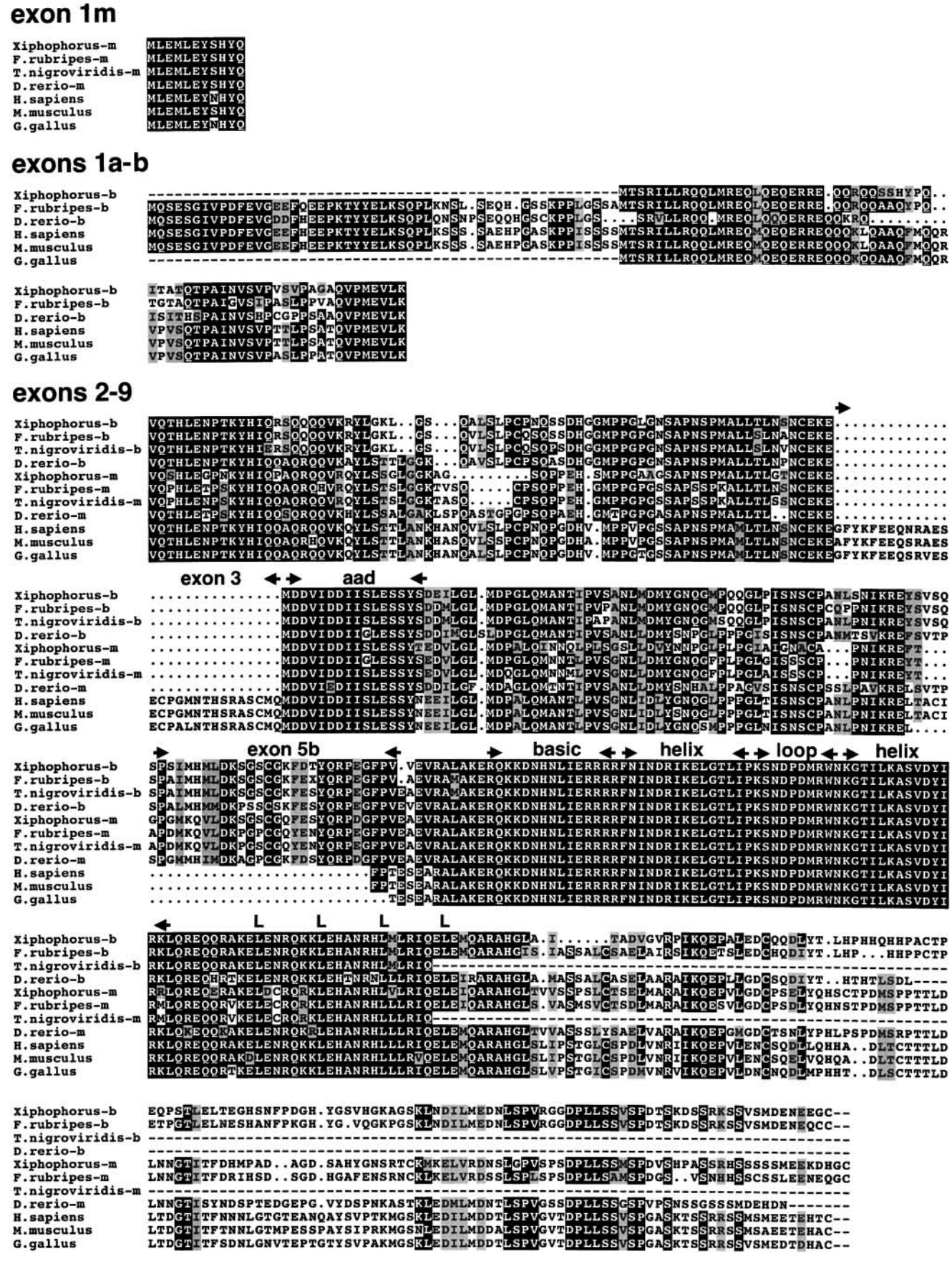

FIGURE 1.—Comparison of vertebrate MITF proteins. The conceptual translation products of fish *mitf* genes are compared to other vertebrate MITFs. Identical residues are white type on black and conservative substitutions are black type on gray (drawn with MacBoxshade). The acidic activation domain (aad), the basic region, the helix-loop-helix domain, and the leucine residues (L) structuring the Zip region are shown. Accession numbers: MITF-m zebrafish, AAD48371; MITF-b zebrafish, AAK95588; MITF mouse, AAF81266; MITF human, I38024; MITF chicken, AAF67467.

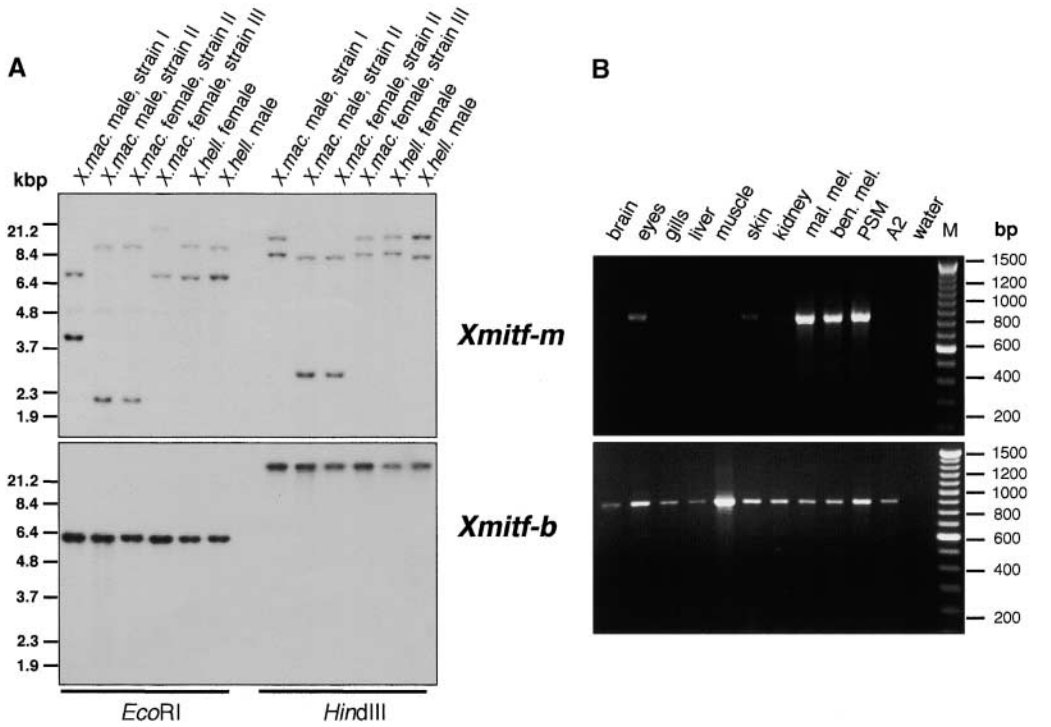

FIGURE 2.—Genomic structure (A) and expression (B) of *mitf* genes in Xiphophorus. (A) Genomic DNAs from several *X. maculatus* and *X. helleri* genotypes (see MATERIALS AND METHODS) were hybridized with X*mitf*-m- and X*mitf*-b-specific probes. The blot was first hybridized with the full-length X*mitf*-m cDNA (top), stripped, and rehybridized with a X*mitf*-b cDNA fragment (bottom). Similar results were also obtained in experiments using the same DNAs digested with *Bam*HI and with DNAs from seven other genotpyes (data not shown). (B) The presence of X*mitf* transcripts was analyzed by RT-PCR, in which primer combinations MiEx-1for/MiEx9rev3 for X*mitf*-m (top) and MiExBfor3/MiEx-9rev3 for X*mitf*-b (bottom) were used for the amplification of specific cDNA of similar length. For the amplification of X*mitf*-b transcripts five times more cDNA was used than for the amplification of X*mitf*-m. RNAs were prepared from normal tissues of backcross hybrids between *X. maculatus* and *X. helleri* and from benign (ben. mel.) and malignant (mal. mel.) melanoma. In addition to tissue RNAs, RNAs from the Xiphophorus melanoma cell line PSM and the embryonic cell line A2 were included in the analysis. M is a DNA size marker and water a control reaction without template.

parisons, the number of synonymous (silent) substitutions was considerably greater than the number of nonsynonymous substitutions (min.–max. of K_s/K_a rate): 18.1–22.7 between *mitf*-b fish sequences, 7.6–20.4 between *mitf*-m fish sequences, 8.5–23.2 between *mitf*-b and *mitf*-m fish sequences, 13.2–32.6 between *mitf*-b fish sequences and sequences from higher vertebrates, 11.6–28.4 between *mitf*-m fish sequences and sequences from higher vertebrates, and >40 between *mitf* sequences from higher vertebrates. This strongly suggests that all

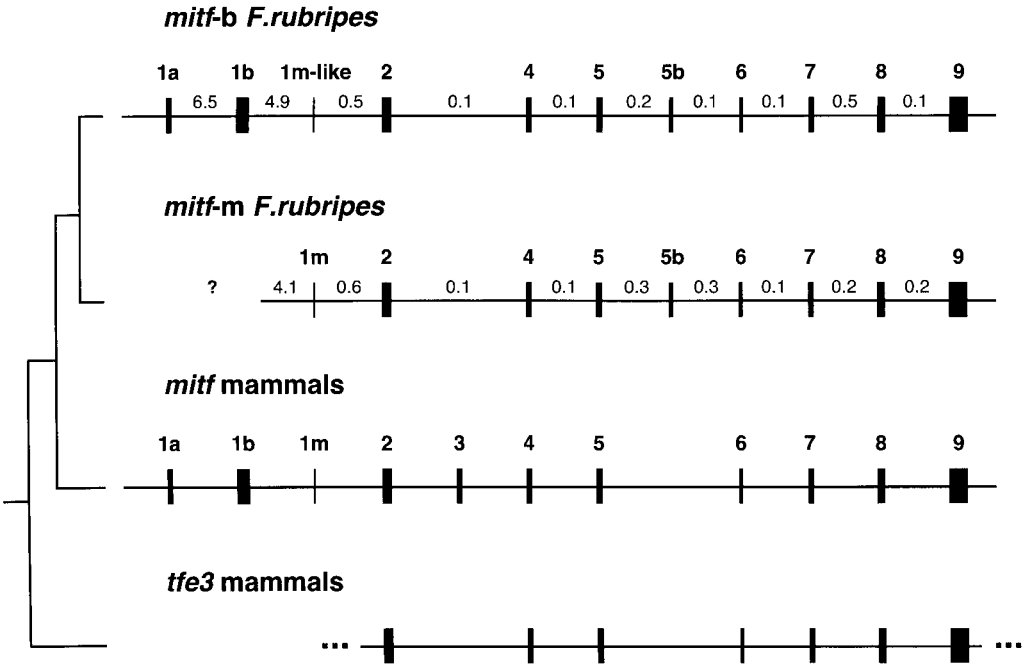

FIGURE 3.—Genomic structure of Fugu and mammalian *mitf* genes. Exons are numbered according to the homologous regions of the human *mitf* gene. The approximate size of Fugu introns is given in kilobases.

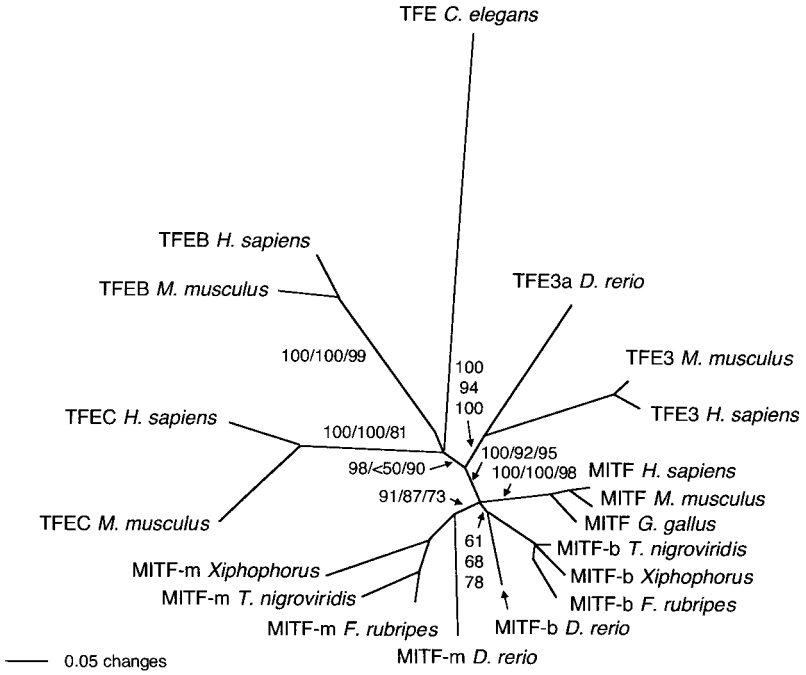

FIGURE 4.—Phylogenetic analysis of MITF-related proteins. Protein regions encoded by exons 1a/1b/1m have been removed for analysis. The tree (neighbor joining) is unrooted. Bootstrap values (%) using neighbor joining (1000 replicates, first value) and maximum parsimony analyses (100 replicates, second value), as well as the reliability values (%) for maximum-likelihood analysis (quartet puzzling, 10,000 puzzling steps, third value) are given. Branches with <50% support have been collapsed. Only the maximum parsimony supported the presence of a fish-specific group containing the MITF-b and MITF-m sequences (bootstrap value 86%). Accession numbers: MITFm *D. rerio*, AAD48371; MITFb *D. rerio*, AAK95588; MITF *Mus musculus*, AAF81266; MITF *Homo sapiens*, I38024; MITF *Gallus gallus*, AAF67467; TFE3a *D. rerio*, AAK95589; TFE3 *H. sapiens*, P19532; TFE3 *M. musculus*, Q64092; TFEB *H. sapiens*, CAB54146; TFEB *M. musculus*, NP_035679; TFEC *M. musculus*, AF077742; TFEC *H. sapiens*, D43945; TFE *Caenorhabditis elegans*, AAB37997.

fish *mitf*-b and -m genes are not pseudogenes but on the contrary mostly evolved under purifying selection. The fact that the K_s/K_a ratio seems to be smaller in comparisons involving fish *mitf* sequences than between sequences from higher eukaryotes might be due to relaxed selective constraints after duplication (LYNCH and CONERY 2000).

Differential exon recruitment or degeneration during *mitf* evolution in vertebrates: Comparative analysis of *mitf* cDNA sequences from fish and higher vertebrates revealed that the alternatively spliced bird/mammalian exon 3 was absent from all fish cDNAs described so far (Figure 1). Analysis of the intervening sequence (maximal size 80 nt) between exons 2 and 4 (according to the mammalian nomenclature) in the *mitf*-b and *mitf*-m genes from both Japanese and freshwater pufferfishes showed that this exon is not present in the *mitf* genes (Figure 3). An additional exon (termed 5b) was found in cDNA and genomic sequences in fish but not in mammals (Figures 1 and 3). Neither intron 3 nor intron 5b could be detected in the human *tfe3* gene (Figure 3). This strongly suggests the acquisition of intron 3 and intron 5b in higher vertebrates and fish, respectively, after divergence of the mammalian/bird lineage from the teleost lineage.

According to the phylogenetic analysis, both *F. rubripes* and *T. nigroviridis mitf*-m genes contain a typical 1m exon about 700 nt upstream from exon 2 (Figures 1, 3, and 5). This exon is predicted by half of the gene structure analysis programs available on the NIX server (http://menu.hgmp.mrc.ac.uk/menu-bin/Nix). The 5′ splice site of intron 1 in both *mitf*-m genes (CAG/GTGAGA) is conserved between both pufferfishes and is closely related to the consensus of 5′ canonical splice sites in mammals ($M_{70}A_{60}G_{80}/GTR_{95}A_{71}G_{81}T_{46}$; BURSET *et al.* 2001). Interestingly, 1m-exon-like sequences were also detected in the same region in both *F. rubripes* and *T. nigroviridis mitf*-b genes (Figures 3 and 5). However, these sequences were degenerated. Their putative translation products lack amino acid residues that are highly conserved in MITF-m proteins from fish and higher vertebrates (Figure 5). While the 1m-like sequence of the Fugu *mitf*-b gene was still predicted as an exon by several programs from the NIX package, this sequence was not identified as an exon in the Tetraodon *mitf*-b gene. This might be due to a less favorable splicing context (GAG/GTATGG), in which the A present in 71% of 5′ splice sites at position +4 (BURSET *et al.* 2001) has been replaced by a T, found in only 9.3% of 5′ splicing sites in mammals. Sequences present in the Tetraodon database did not allow the complete reconstruction of the *mitf*-b gene and the identification of exons 1a and 1b, probably because of the relatively large size of introns 1 and 2 as observed in the Fugu *mitf*-b gene (Figure 3). No exons 1a and 1b could be detected within 4.5- and 2.2-kb genomic sequences upstream from the 1m exon of *mitf*-m genes in *F. rubripes* and *T. nigroviridis*, respectively, but the presence of more distant degenerated exon 1a/b sequences in these genes cannot be excluded.

DISCUSSION

Melanocyte development as well as postnatal proliferation and survival is dependent on MITF, the gene product of the *microphthalmia* locus. In mammals and birds a single gene within this locus, *mitf*, encodes several proteins with different N termini. Use of alternative

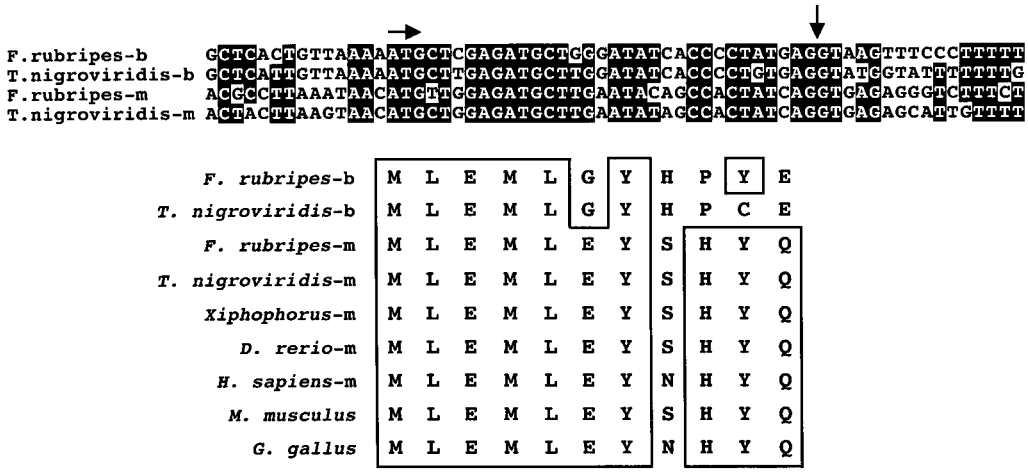

FIGURE 5.—Exon 1m-related sequences in pufferfish *mitf* genes. Horizontal arrow shows the putative ATG start codon, vertical arrow shows the position of the predicted 5′ splice site of intron 1 in *mitf*-m. Amino acid residues that are identical within the sequence encoded by all *bona fide* m exons in vertebrates are boxed.

promoters and first exons results in different mRNAs encoding these MITF isoforms. One of them, MITF-m, is of critical importance for the neural crest-derived melanocytic cell lineage. In addition to its function in normal pigmentation, MITF-m might also play a role in transformed pigment cells, as it has been detected in melanomas in contrast to many other melanocytic markers, which are lost during the transformation process (KING *et al.* 1999).

In contrast to the situation in higher vertebrates, we have now demonstrated the presence of at least two distinct *mitf* genes in the small aquarium fish Xiphophorus, an animal model for melanoma formation with well-established tumor genetics (SCHARTL 1995), and in both pufferfishes *F. rubripes* and *T. nigroviridis*. These results, added to the recent report of the presence of two *mitf* genes in zebrafish (LISTER *et al.* 2001), show the presence of duplicate *mitf* genes in the teleost fish lineage for at least 100–120 million years (minimal age of divergence between the zebrafish and the Xiphophorus/pufferfish lineages). This is supported by the fact that fish MITF-m and MITF-b proteins form two distinct monophyletic groups. According to the analysis of K_s/K_a ratios in four different fish species, the duplicate *mitf* genes are not pseudogenes but rather evolved under purifying selection.

While the phylogenetic analysis using maximum parsimony supports a duplication of the *mitf* genes within the fish lineage after its separation from the bird/mammalian lineage, the neighbor joining and maximum-likelihood methods did not allow resolving the relative position of both fish MITF protein groups. A hypothesis alternative to fish-specific duplication would imply that the *mitf* gene was duplicated before the separation between the bird/mammalian and fish lineages. While both copies would have been maintained in teleosts, one duplicate would have been lost in higher vertebrates. The remaining copy should be then more related to either the *mitf*-b or the *mitf*-m fish gene.

Comparative structural analysis of vertebrate *mitf* cDNA and genomic sequences showed that intron 3 (according to the mammalian nomenclature) is present in birds and mammals but not in teleosts, while intron 5b is found in both *mitf*-b and *mitf*-m genes in fish but not in the unique *mitf* gene of higher vertebrates. Since both introns are absent from the *mitf*-related *tfe3* genes, they have probably been recruited during evolution of the *mitf* genes in vertebrates. If we consider the scenario of duplication of *mitf* before separation of fish and birds/mammals, intron 5b should have been present in *mitf* before duplication (since intron 5b is present in both *mitf*-m and *mitf*-b in fish) and should have been lost in the remaining copy of higher vertebrates. On the other hand, if we consider the hypothesis of a fish-specific duplication of *mitf*, intron 5b might have been recruited within the fish lineage before the duplication event.

The fish-specific duplication hypothesis, suggested by the maximum parsimony analysis, implies only one event of gene duplication. Although we cannot exclude that the *mitf* gene was duplicated before the separation between fish and birds/mammals, this alternative hypothesis would involve additional events (loss of one *mitf* duplicate and loss of intron 5b in the remaining copy in higher vertebrates). Hence, the fish-specific duplication hypothesis appears to be the most parsimonious one. Absence of strong support for this hypothesis in the neighbor joining and maximum-likelihood analyses might be explained by modified evolutionary rates due to relaxed selective constraints after duplication, as suggested by the analysis of the K_s/K_a ratios.

The fish-specific duplication hypothesis fits accumulating evidence that during the course of evolution teleost fishes underwent a whole genome duplication since they separated from the last common ancestor that they shared with amphibians, reptiles, birds, and mammals (AMORES *et al.* 1998; GATES *et al.* 1999; MEYER and SCHARTL 1999; TAYLOR *et al.* 2001; VAN DE PEER *et al.*

2001; but see Robinson-Rechavi *et al.* 2001). As a consequence in many instances fish have two similar genes where mammals have only one. This genome duplication might have arisen as early as 300–450 million years ago (Taylor *et al.* 2001; Van de Peer *et al.* 2001). The divergence between *mitf*-m and *mitf*-b nucleic acid and protein sequences, even at synonymous sites, is always higher than the divergence between chicken and mammal sequences (divergence ~300 million years ago). This is consistent with an ancient event of duplication of the *mitf* genes. It has been estimated that ~20% of the duplicate genes, particularly genes encoding DNA-binding proteins (like MITF), have been maintained in teleost genomes by neofunctionalization of a duplicate and/or by subfunctionalization (Force *et al.* 1999; Postlethwait *et al.* 2000; Van de Peer *et al.* 2001). This last phenomenon could arise as a consequence of the differential decay of specific sequences in each gene copy and the subsequent need for the presence of both duplicates with complementary activities to achieve the function(s) of the original unique gene (Force *et al.* 1999). We present evidence for a subfunctionalization where two alternative transcripts of the same gene with different functions have been replaced by two distinct genes. This is associated with the degeneration of at least one alternative exon: the 1m exon specific for the m-form of MITF in higher vertebrates shows a degenerated coding sequence in the *mitf*-b gene of both pufferfishes. This might be associated, at least in Tetraodon, with the degeneration of the flanking splicing site. As observed in other cases of subfunctionalization of duplicate fish genes reported to date (for examples, see Chiang *et al.* 2001a,b; Serluca *et al.* 2001), fish *mitf*-m and *mitf*-b show different expression patterns during embryonic development in zebrafish (Lister *et al.* 2001) and in adult tissues in Xiphophorus. Hence, subfunctionalization of *mitf* genes in fish is also probably associated with the degeneration of specific promoters and/or regulatory sequences. The order of degenerative events remains to be determined: mutations may have first altered the coding sequence and/or the splicing site of an alternative exon in duplicate A (for example exon 1m in *mitf*-b) but not in duplicate B. In this way the selection for functionality on the promoter driving the expression of the transcript containing the alternative exon in duplicate A would have been abolished. Alternatively, alteration of promoter or specific regulatory sequences necessary for the expression of an alternative transcript in duplicate A but not in duplicate B would allow accumulation in duplicate A of nonconservative mutations in sequences that are specific for this alternative transcript.

We cannot exclude that the duplication of the *mitf* genes was a single gene duplication event that happened very early in the teleost lineage. More data are needed from other fish genomes to understand the genomic mechanism and the evolutionary significance underlying the duplication of the *mitf* gene in fish genomes.

We thank U. Hornung and E. Geissinger for RNAs and Jim A. Lister for sharing the zebrafish *nacre* sequence with us prior to publication. We are grateful to Cécile Fischer, Alain Bernot, Jean Weissenbach, and the other members of the Tetraodon Genome project (Genoscope, Evry, France), as well as to the International Fugu Sequencing Consortium, for making sequencing data available to the public. Fugu data have been provided freely by the Fugu Genome Consortium for use in this publication only. This work was supported by grants to M.S. supplied by the Deutsche Forschungsgemeinschaft through SFB 465 ("Entwicklung und Manipulation pluripotenter Zellen") and the Fonds der Chemischen Industrie. J.-N.V. is supported by the BioFuture program of the German Bundesministerium für Bildung und Forschung (BMBF).

LITERATURE CITED

Amae, S., N. Fuse, K. Yasumoto, S. Sato, I. Yajima *et al.*, 1998 Identification of a novel isoform of microphthalmia-associated transcription factor that is enriched in retinal pigment epithelium. Biochem. Biophys. Res. Commun. **24:** 710–715.

Amores, A., A. Force, Y. L. Yan, L. Joly, C. Amemiya *et al.*, 1998 Zebrafish *hox* clusters and vertebrate genome evolution. Science **282:** 1711–1714.

Baudler, M., J. Duschl, C. Winkler, M. Schartl and J. Altschmied, 1997 Activation of transcription of the melanoma inducing *Xmrk* oncogene by a GC-box element. J. Biol. Chem. **272:** 131–137.

Burset, M., I. A. Seledtsov and V. V. Solovyev, 2001 SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res. **29:** 255–259.

Chiang, E. F., C. I. Pai, M. Wyatt, Y. L. Yan, J. Postlethwait *et al.*, 2001a Two *sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. Dev. Biol. **231:** 149–163.

Chiang, E. F., Y. L. Yan, Y. Guiguen, J. Postlethwait and B. Chung, 2001b Two Cyp19 (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. Mol. Biol. Evol. **18:** 542–550.

Crollius, H. R., O. Jaillon, C. Dasilva, C. Ozouf-Costaz, C. Fizames *et al.*, 2000 Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. Genome Res. **10:** 939–949.

Elgar, G., R. Sandford, S. Aparicio, A. Macrae, B. Venkatesh *et al.*, 1996 Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). Trends Genet. **12:** 145–150.

Elgar, G., M. S. Clark, S. Meek, S. Smith, S. Warner *et al.*, 1999 Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. Genome Res. **9:** 960–971.

Fischer, C., C. Ozouf-Costaz, H. Roest Crollius, C. Dasilva, O. Jaillon *et al.*, 2000 Karyotype and chromosome location of characteristic tandem repeats in the pufferfish *Tetraodon nigroviridis*. Cytogenet. Cell Genet. **88:** 50–55.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Gates, M. A., L. Kim, E. S. Egan, T. Cardozo, H. I. Sirotkin *et al.*, 1999 A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. Genome Res. **9:** 334–347.

Hallsson, J. H., J. Favor, C. Hodgkinson, T. Glaser, M. L. Lamoreux *et al.*, 2000 Genomic, transcriptional and mutational analysis of the mouse microphthalmia locus. Genetics **155:** 291–300.

Hodgkinson, C. A., K. J. Moore, A. Nakayama, E. Steingrimsson, N. G. Copeland *et al.*, 1993 Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. Cell **74:** 395–404.

Hughes, M. J., J. B. Lingrel, J. M. Krakowsky and K. P. Anderson, 1993 A helix-loop-helix transcription factor-like gene is located at the mi locus. J. Biol. Chem. **268:** 20687–20690.

Israel, D. I., 1993 A PCR-based method for high stringency screening of DNA libraries. Nucleic Acids Res. **21:** 2627–2631.

Kallman, K. D., 1975 The platyfish, *Xiphophorus maculatus*, pp. 81–132 in *Vertebrates of Genetic Interest*, edited by R. C. King. Plenum Press, New York.

King, R., K. N. Weilbaecher, G. McGill, E. Cooley, M. Mihm *et al.*, 1999 Microphthalmia transcription factor. A sensitive and specific melanocyte marker for melanoma diagnosis. Am. J. Pathol. **155:** 731–738.

Kuhn, C., U. Vielkind and F. Anders, 1979 Cell cultures derived from embryos and melanoma of poeciliid fish. In Vitro **15:** 537–544.

Lister, J. A., C. P. Robertson, T. Lepage, S. L. Johnson and D. W. Raible, 1999 nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. Development **126:** 3757–3767.

Lister, J. A., J. Close and D. W. Raible, 2001 Duplicate *mitf* genes in zebrafish: complementary expression and conservation of melanogenic potential. Dev. Biol. **237:** 333–344.

Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155.

Meyer, A., and M. Schartl, 1999 Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr. Opin. Cell Biol. **11:** 699–704.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, New York.

Postlethwait, J. H., I. G. Woods, P. Ngo-Hazelett, Y. L. Yan, P. D. Kelly *et al.*, 2000 Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res. **10:** 1890–1902.

Rehli, M., N. Den Elzen, A. I. Cassady, M. C. Ostrowski and D. A. Hume, 1999 Cloning and characterization of the murine genes for bHLH-ZIP transcription factors TFEC and TFEB reveal a common gene organization for all MiT subfamily members. Genomics **56:** 111–120.

Robinson-Rechavi, M., O. Marchand, H. Escriva and V. Laudet, 2001 An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr. Biol. **11:** 458–459.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Schartl, M., 1995 Platyfish and swordtails: a genetic system for the analysis of molecular mechanisms in tumor formation. Trends Genet. **11:** 185–189.

Schartl, M., B. Wilde and J. Parzefall, 1995 Evolutionary origin of a parthenoform, the Amazon Molly *Poecilia formosa*, on the basis of a molecular genealogy. Evolution **49:** 827–835.

Serluca, F. C., A. Sidow, J. D. Mably and M. C. Fishman, 2001 Partitioning of tissue expression accompanies multiple duplications of the Na+/K+ ATPase alpha subunit gene. Genome Res. **11:** 1625–1631.

Smith, S. D., P. M. Kelley, J. B. Kenyon and D. Hoover, 2000 Tietz syndrome (hypopigmentation/deafness) caused by mutation of MITF. J. Med. Genet. **37:** 446–448.

Steingrimsson, E., K. J. Moore, M. L. Lamoreux, A. R. Ferre-D'Amare, S. K. Burley *et al.*, 1994 Molecular basis of mouse *microphthalmia* (*mi*) mutations helps explain their developmental and phenotypic consequences. Nat. Genet. **8:** 256–263.

Strimmer, K., and A. von Haeseler, 1996 Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:** 964–969.

Swofford, D. L., 1989 *PAUP: Phylogenetic Analysis Using Parsimony*. Smithsonian Institution, Washington, DC.

Takeda, K., C. Takemoto, I. Kobayashi, A. Watanabe, Y. Nobukuni *et al.*, 2000 Ser298 of MITF, a mutation site in Waardenburg syndrome type 2, is a phosphorylation site with functional significance. Hum. Mol. Genet. **9:** 125–132.

Tassabehji, M., V. E. Newton and A. P. Read, 1994 Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. Nat. Genet. **8:** 251–255.

Taylor, J. S., Y. Van de Peer, I. Braasch and A. Meyer, 2001 Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos. Trans. R. Soc. Lond. B Biol. Sci. **356:** 1661–1679.

Udono, T., K. Yasumoto, K. Takeda, S. Amae, K. Watanabe *et al.*, 2000 Structural organization of the human microphthalmia-associated transcription factor gene containing four alternative promoters. Biochim. Biophys. Acta **1491:** 205–219.

Vachtenheim, J., and H. Novotna, 1999 Expression of genes for microphthalmia isoforms, Pax3 and MSG1, in human melanomas. Cell. Mol. Biol. **45:** 1075–1082.

Van de Peer, Y., J. S. Taylor, I. Braasch and A. Meyer, 2001 The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J. Mol. Evol. **53:** 436–446.

Wakamatsu, Y., 1981 Establishment of a cell line from the platyfish-swordtail hybrid melanoma. Cancer Res. **41:** 679–680.

Wittbrodt, J., A. Meyer and M. Schartl, 1998 More genes in fish? Bioessays **20:** 511–515.

Yajima, I., S. Sato, T. Kimura, K. Yasumoto, S. Shibahara *et al.*, 1999 An *L1* element intronic insertion in the black-eyed white (Mitf[mi-bw]) gene: the loss of a single Mitf isoform responsible for the pigmentary defect and inner ear deafness. Hum. Mol. Genet. **8:** 1431–1441.

Yasumoto, K., S. Amae, T. Udono, N. Fuse, K. Takeda *et al.*, 1998 A big gene linked to small eyes encodes multiple Mitf isoforms: many promoters make light work. Pigm. Cell Res. **11:** 329–336.

Communicating editor: D. J. Grunwald