Evolution of Duplicated *reggie* Genes in Zebrafish and Goldfish

Edward Málaga-Trillo, Ute Laessing, Dirk M. Lang, Axel Meyer, Claudia A.O. Stuermer

Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Abstract. Invertebrates, tetrapod vertebrates, and fish might be expected to differ in their number of gene copies, possibly due the occurrence of genome duplication events during animal evolution. *Reggie* (*flotillin*) genes code for membrane-associated proteins involved in growth signaling in developing and regenerating axons. Until now, there appeared to be only two *reggie* genes in fruitflies, mammals, and fish. The aim of this research was to search for additional copies of *reggie* genes in fishes, since a genome duplication might have increased the gene copy number in this group. We report the presence of up to four distinct *reggie* genes (two *reggie-1* and two *reggie-2* genes) in the genomes of zebrafish and goldfish. Phylogenetic analyses show that the zebrafish and goldfish sequence pairs are orthologous, and that the additional copies could have arisen through a genome duplication in a common ancestor of bony fish. The presence of novel *reggie* mRNAs in fish embryos indicates that the newly discovered gene copies are transcribed and possibly expressed in the developing and regenerating nervous system. The intron/exon boundaries of the new fish genes characterized here correspond with those of human genes, both in location and phase. An evolutionary scenario for the evolution of *reggie* intron-exon structure, where loss of introns appears to be a distinctive trait in invertebrate *reggie* genes, is presented.

Key words: *reggie* — *flotillin* — Gene duplication — Microdomains — Axon growth — Axon regeneration — *Carassius auratus* — *Danio rerio*

Correspondence to: E. Málaga-Trillo; *email:* edward.malaga@uni-konstanz.de

Introduction

Reggie-1 and reggie-2 (homologous to flotillin-2 and flotillin-1, see below) are two closely related 48 kD intracellular and membrane-associated proteins whose encoding genes have been identified so far in goldfish (Schulte et al. 1997), rats (Lang et al. 1998), mice (Cho et al. 1995; Bickel et al. 1997), humans (Schroeder et al. 1994), and fruitflies (Galbiati et al. 1998). Reggies are co-expressed in many cell types and define specific plasma membrane microdomains (Lang et al. 1998) In contrast to what has been suggested by others (Bickel et al. 1997; Volonte et al. 1999), they do not occur in caveolae—small plasma membrane microdomains characterized by the expression of caveolin—but are expressed clearly outside of these structures (Lang et al. 1998). Moreover, in neurons and T-lymphocytes—which do not possess caveolae—reggie proteins occur in their own specific microdomains (Lang et al. 1998). In goldfish and rats, reggie-1 and -2 are upregulated during axon growth and regeneration of retinal ganglion cells (RGC) (Schulte et al. 1997; Lang et al. 1998). Although the precise nature of their role in the central nervous system is not fully understood, two lines of evidence strongly suggest that reggie-1 and -2 participate in molecular signaling across the plasmamembrane: 1) In cultured neurons, reggie-1 and -2 occur all along axons and are associated with GPI-linked CAMs and fyn tyrosine kinase (Stuermer et al. 2001); 2) In activated T-lymphocytes, reggie-1 and -2 occur in micropatches along the plasma membrane and are recruited into the T-cell receptor signaling complex (along with Thy-1 and GM-1) (Lang et al. 1998; Stuermer et al. 2001).

It has long been suggested that two rounds of genome duplication facilitated the appearance of novel morphologies early in vertebrate evolution by increasing the

levels of genetic complexity (Ohno 1970; Ohno 1999). This hypothesis would explain why vertebrates can have up to four copies of genes that are present as single copies in invertebrates. The recent discovery in the zebrafish of extra copies of several gene families uncovered an additional genome duplication that probably took place in the common ancestor of all teleost fish (Amores et al. 1998; Wittbrodt et al. 1998; Meyer and Schartl 1999; Meyer and Málaga-Trillo 1999). The generation of new gene copies through genome duplications raises questions about how much functional redundancy can be tolerated in genomes and whether duplicated copies are randomly deleted, inactivated, or selected to perform a novel function (Málaga-Trillo and Meyer 2001). To answer these questions, it is necessary to identify as many families of duplicated genes as possible, and to examine the evolution of their functions.

So far only one copy of each *reggie-*(*flotillin*) gene has been characterized in fruitflies, mammals, and goldfish, and yet, if the additional fish genome-duplication hypothesis is correct, then for any given gene family, fish are expected to have up to twice as many copies as mammals and up to eight times as many copies as invertebrates, although average gene copy ratios between these groups cannot always be expected to remain whole numbers because not all duplicated genes are maintained for a long evolutionary time. Some duplicated copies may be lost as pseudogenes or rendered unrecognizable as "junk DNA." It is therefore possible that: 1) Only one functional *reggie-1* or *-2* gene was retained in vertebrate genomes, or that, 2) Additional *reggie* copies exist (or existed) in vertebrates but remain undiscovered. Thus, *reggie* genes offer a case study for the evolution of genetic redundancy and the functional fate of duplicated gene copies. Given their importance in the development of the vertebrate nervous system (Schulte et al. 1997; Lang et al. 1998), it is of interest to know how duplication of these genes affected their function. Until now, the only fish *reggie* genes known are those of the goldfish *Carassius auratus* (Lang et al. 1998), a model system routinely employed for the study of retinotectal development and regeneration of retinal axons (Gaze and Sharma 1970). Despite their obvious biomedical interest, *reggie* genes have not yet been characterized in the zebrafish, *Danio rerio,* the system of choice for the molecular characterization of genes important for diverse aspects of vertebrate development (Haffter et al. 1996). In order to evaluate new evidence for the fish genome duplication hypothesis, and as a first step towards understanding the functional evolution of redundant gene products in the nervous system, we set out to characterize the *reggie* genes of zebrafish and goldfish.

Materials and Methods

Nucleic acid isolation. Zebrafish and goldfish total genomic DNAs were prepared from fresh muscle tissue of homozygous individuals using ATL Lysis Buffer (Qiagen, Hilden). We followed standard methods (Sambrook et al. 1989) for the purification steps.

Synthesis of zebrafish 1st strand cDNA for RT-PCR. Total RNA was isolated from two and three day old zebrafish embryos by the RNAzol B method (WAK Chemie, Bad Soden). First strand cDNA was generated using Oligo-dT primers and the Superscript Reverse Transcriptase (Gibco BRL, Eggenstein) or the Ready-to-go T-primed First-strand Synthesis Kit (Amersham Pharmacia), as outlined by the manufacturer.

Preparation of goldfish cDNA libraries. Two weeks after optic nerve transection, mRNA from goldfish retinae was prepared with the Quick Prep mRNA Purification Kit (Amersham Pharmacia), and reverse transcribed using Superscript RT (Gibco BRL). After second strand synthesis and ligation to *Eco*RI adaptors (Amersham Pharmacia), the cDNAs were cloned into Lambda ZapII (Stratagene, Heidelberg). The titer of the cDNA library was amplified from 2.4×10^6 PFU/ml to 4.5×10^{10} PFU/ml.

Polymerase chain reaction (PCR). Reggie-1 and *-2* genes fragments were amplified with several pairs of degenerate primers located at various points along the *reggie-1* and *-2* coding sequence (Table 1). One µl of template was added to 19 µl PCR reaction mix containing 0.5 units of Red Taq® Polymerase (Sigma-Aldrich, Taufkirchen), 1× Red Taq Buffer, 800 µM dNTPs, and 0.5 µM primers. PCR reactions were performed in GeneAmp® PCR System 9700 thermocyclers (PE Applied Biosystems, Weiterstadt) using the following conditions: Initial denaturation for 2 min at 94°C, followed by 35–40 cycles of: denaturation at 94°C for 1 min, annealing at 50–56°C for 1 min, extension at 72°C for 2 min; and a final extension step at 72°C for 10 min.

Cloning of PCR products. Amplification products were directly used for cloning, or first excised from 2% agarose gels (Gibco BRL). After purification with the Qiaquick PCR Purification Kit (Qiagen), the DNAs were ligated into the Topo II vector (Invitrogen, Karlsruhe) and the ligation mixes used to transform *E. coli.* Transformations were plated onto LB-agar plates containing the appropriate amounts of ampicillin, X-gal, and IPTG (Sambrook et al. 1989). White recombinant bacteria were grown and tested by PCR for the presence of the correct insert.

DNA sequencing. PCR products were purified with the aid of the QIAquick PCR purification kit (Qiagen), and plasmid templates from cloned amplification products were isolated using the QIAprep Spin Miniprep Kit (Qiagen). Sequencing reactions were set up in 10 µl volumes containing up to 7 µl of DNA, 1 µl of 10 µM primer and 2 µl of Big Dye Terminator Cycle Sequencing Ready Reaction Mix (DNA Sequencing Kit, PE Applied Biosystems). The amount of template used for cycle sequencing and the thermal cycling profile were chosen according to the kit's guidelines. Unincorporated dye terminators were removed by EtOH/NaOAc precipitation and the purified reactions were loaded onto 0.2 mm-thick 5% Long Ranger™ gels and ran on an ABI 377 automated DNA sequencer.

DNA sequence analysis and drawing of dendrograms. DNA sequences were edited and analyzed using PE Applied Biosystems software (Sequence Analysis and Sequence Navigator) as well as the GCG Wisconsin Package Version 9.1 (The Genetics Computer Group 1997). Final sequences have been deposited in the GenBank database under the accession codes AF315945, AF315946, AF315947, AF315948, and AF315949. Sequence alignments were made with DNA Star's MegAlign, and phylogenetic analysis was performed using both distance and maximum parsimony algorithms with Paup 4.0b4 (Swofford 2000). Previously characterized *reggie* sequences were downloaded
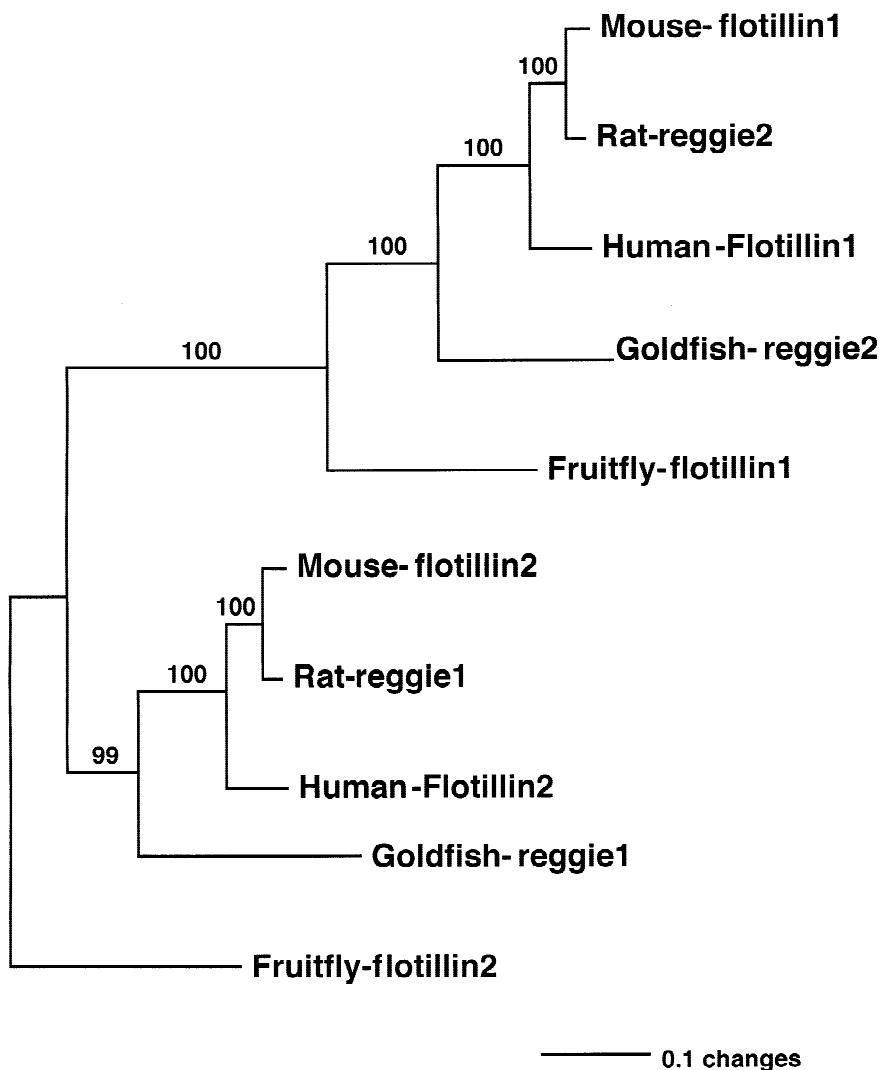
Fig. 1. Neighbor-joining tree of previously characterized *reggie* (Reg) and *flotillin* (Flo) and DNA sequences. An alignment of full-length coding sequences was used as a basis for phylogenetic reconstruction. The distance matrix was generated using the Kimura two-parameter method. The horizontal length of each branch is proportional to the estimated genetic distance (number of substitutions) between taxa. Bootstrap values (1000 replications) are indicated on each branch.

from Genbank. Accession numbers for the *flotillin-1/reggie-2* sequences of Drosophila, goldfish, mouse, rat, and human are AF044734 (genomic: AE003810), U33556, NM_008027, U60976, and NM-005803 (genomic: NT_001520). Accession numbers for the *flotillin-2/reggie-1* sequences of Drosophila, goldfish, mouse, rat, and human are AF044916 (genomic: AE003497), L36867, NM_008028, AF023303, and NM_004475 (genomic: AC024267). It should be noted that several *reggie 1* and *2* cDNA sequences have been reported in the rat. This does not imply that the rat has more *reggie* genes. Based on comparison with human genomic sequences (not shown), the additional rat sequences are rather the result of the isolation of heterogeneous nuclear RNAs and alternative splicing forms. For instance, in the case of *reggie 1,* a splice variant sequence starts at exon 4 and another one at exon 5.

For distance analysis, genetic distances were estimated with the Kimura two-parameter method (Kimura 1980). Dendrograms were prepared using the neighbor-joining (NJ) tree constructing method (Saitou and Nei 1987) and reliability of the tree topologies was evaluated through 1000 bootstrap replications (Felsenstein 1985).

Results and Discussion

Homology of reggie and flotillin Genes

The *reggie* genes of goldfish and rat (Schulte et al. 1997; Lang et al. 1998) were reported independently of the

flotillin genes of fruitflies, humans, and mice (Schroeder et al. 1994; Cho et al. 1995; Galbiati et al. 1998). Nevertheless, their highly similar DNA sequences and expression patterns strongly suggest that these genes are closely related and members of the same gene family (above references and unpublished Medline records). We performed phylogenetic analysis in order to clarify their evolutionary relationships. The neighbor-joining tree in Fig. 1 defines two major sequence homology groups, one containing *reggie-1* and *flotillin-2* sequences, and the other one including *reggie-2* and *flotillin-1* sequences. The same topology was obtained with high bootstrap support when the maximum parsimony method was employed (not shown). Therefore, the current nomenclature is in need of revision: the names "*reggie*" and "*flotillin*" designate homologous genes, but numbering of the loci (1 or 2) is misleading. Here, we will use the names "*reggie-1*" and "*reggie-2*" to refer to *reggie-1/flotillin-2* and *reggie-2/flotillin-1* groups, respectively The presence of pairs of homologous *reggie* genes in mammals and arthropods indicates that *reggie-1* and *-2* arose before the evolutionary split between protostomes and deu-

Table 1. PCR Primers used in this study

Primer ID	Sequence
1. Reg-19-43F	ACTKKYGGMCCMAAYGARGCVMTSG
2. Reg-193-215F	GGDGTVSCYWTYWCYGTSACHGG
3. Reg-193-215R	CCDGTSACRGWRAWRGSBACHCC
4. Reg-271-288F	SAGMWGTTYCTGGGRAAG
5. Reg-271-288R	CTTYCCCAGRAACWKCTS
6. Reg-360-384F	GACKGTGGAGSAGATYTAYMAGGAC
7. Reg-271-288R	GTCCTKRTARATCTSCTCCACMGTC
8. Reg-568-590F	GRGATGCNGGGATYMGGGARGC
9. Reg-568-590R	GCYTCCCKRATCCCNGCATCYC
10. Reg-720-743F	SMTGGCCTATCAGCTMCAGGYRGC
11. Reg-720-743R	GCYRCCTGKAGCTGATAGGCCAKS
12. Reg-862-885F	GTSMRSMRDCCWGCBGADGCMGAG
13. Reg-862-885R	CTCKGCHTCVGCWGGHYKSYKSAC
14. Reg-1015-1039F	GAGGCYGAGCRRATGRSSMWGAAGG
15. Reg-1015-1039R	CCTTCWKSSYCATYYGCTCRGCCTC
16. Reg-1098-1119R	GAYCTYYYCWGCMAYCTKGGGC
17. GF-Reg1F	GARGTIGCIGCICCIGAYGT
18. GF-Reg1R	GCYTCIGCYTCICCDATYTT
19. GF-Reg2F	AAYGARGCIATGGTIGTITC
20. GF-Reg2R	TCYTCDATYTGYTGYTTIGTYTT
21. Dare-Reg1aXF	ACAGGGATCAGTTTGCCAAG
22. Dare-Reg1aXR	CTCTCTCTGCTTCAGCCACA
23. Dare-Re2aXF	CGATGGTGGTGTCAGGTAAG
24. Dare-Reg2aXR	GGACCCCATGACGTGTGTA
25. Dare-Reg2bXF	GCTCCTCCTCTCATGATTGC
26. Dare-Reg2bXR	GCCTTCTTGGTGTTGACCTC

Numbering corresponds to that of Figure 2. Primer orientation is specified by "F" or "R" at the end of the primer's name.

terostomes. On average, the intraspecific aminoacid sequence similarity between *reggie-1* and *-2* genes is approximately 52.5%, while the similarity between vertebrate and invertebrate *reggies* of the *-1* or *-2* groups is approximately 62.8%. The length of the *reggie* coding region is 429 aminoacids (aa) and appears to be conserved during evolution, although minor length variation can be observed at the 3′ end of some *reggie-2* genes. In all pairwise comparisons, substitutions are randomly distributed along the gene and no mutational hotspots or highly conserved domains are apparent.

Isolation of Novel reggie *Sequences From Fish*

The goldfish and zebrafish sequences reported here were obtained by PCR amplification of genomic and cDNA templates using several combinations of degenerate primers. An alignment of all previously available *reggie-1* and *-2* DNA sequences (not shown) was used as the basis for primer design. In order to increase the specificity of our amplifications towards goldfish/zebrafish sequences, two of the primer pairs (GF-Reg1F/R and GF-Reg2F/R) were constructed specifically for goldfish *reggie-1* and *2* sequences, and used in combination with the degenerate ones. At the time we started with these experiments, no information about the intron-exon organization of *reggie* genes was available. This presented a potential problem for the amplification of genomic targets. Therefore, 20 forward and reverse primers at 13
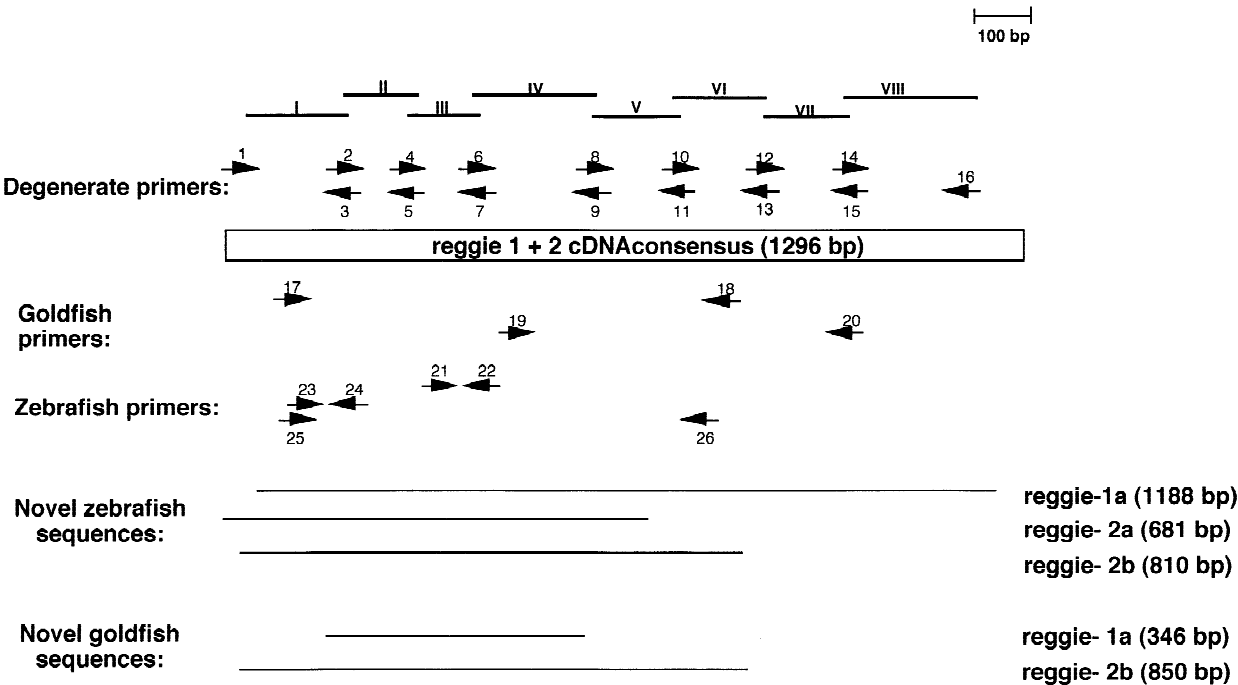
Fig. 2. PCR strategy for the amplification of *reggie* genes in zebrafish and goldfish. Roman numbers indicate the arbitrary amplification targets defined by the position of our degenerate primers. Primers are represented by arrows and numbered according to Table 1. Primer orientation is indicated by the arrows. The final length of the amplification products obtained is shown for each of the novel genes identified in this study.

A. Genome duplication preceeding speciation
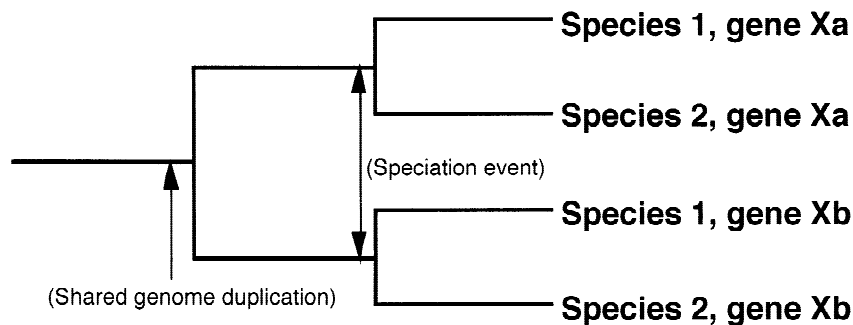
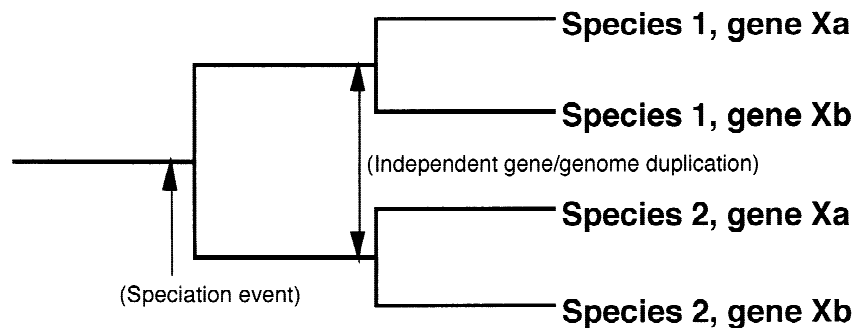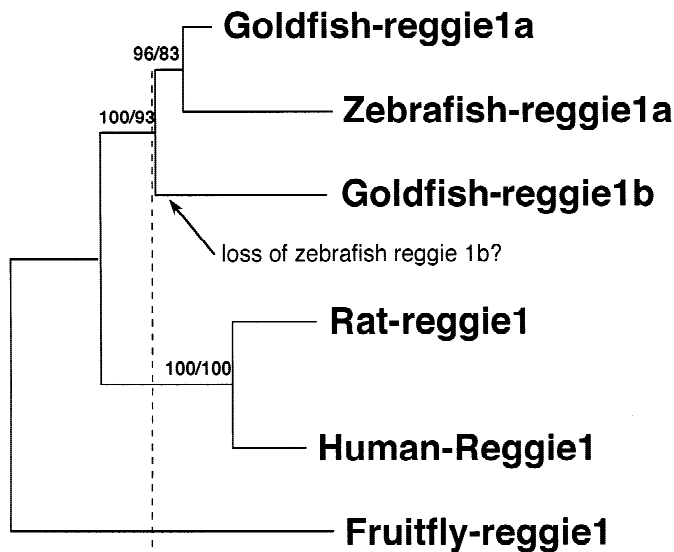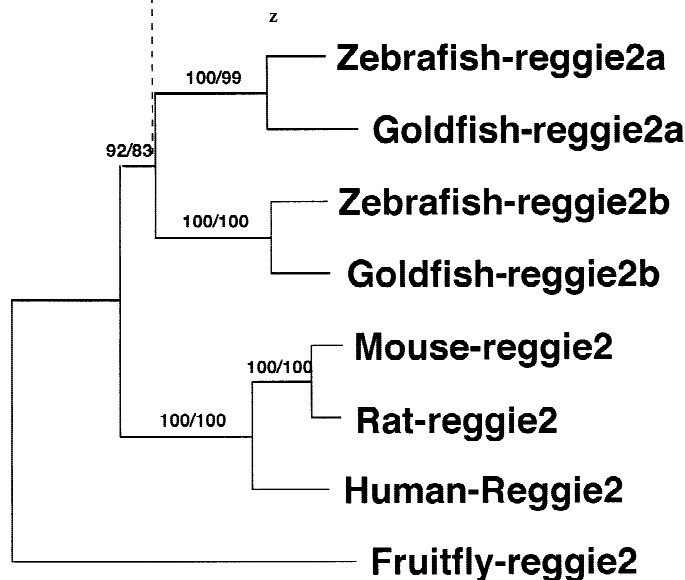## B. Speciation preceeding genome duplication

Fig. 3. Two phylogenetic scenarios for the tree topology of duplicated genes X in hypothetical species 1 and 2, depending on the relative timing of their duplication. **A:** A shared gene duplication before speciation results in trans-specific groupings. **B:** Independent gene duplications (after speciation) result in grouping of genes according to species.

different locations in the least variable segments of the *reggie-1* and *-2* coding region were designed (Table 1). The positions of the PCR primers divide most of the *reggie-1* and *-2* consensus sequence into eight amplification targets (Fig. 2). All primer combinations encompassing one, two, or up to three of these segments were tested. The efficiency of the amplifications varied largely depending on the primer pair and the type of template used. The strongest and cleanest amplification products of expected sizes were cloned and sequenced. Primer combinations 271-288F/360-384R, 360-384F/568-590R, and GF-Reg1F/568-590R amplified *reggie-1* products; primer combinations GF-Reg2F/193-215R, 360-384F/GF-Reg2R, 720-742F/862-885R, and GF-Reg2F/GF-Reg2R amplified *reggie-2* products. Alignment of the amplified sequences identified four and three different *reggie* genes in goldfish and zebrafish, respectively. Two of the goldfish sequences had been previously reported (Schulte et al. 1997), but the other two represent novel *reggie-1* and *-2* genes. The three zebrafish sequences are novel, one of them homologous to *reggie-1* and the other two homologous to *reggie-2* (Fig. 4). These new sequences represent distinct loci and not alleles because they were isolated from single homozygous fish. We designated the four fish genes, *reggie-1a, -1b, -2a,* and *-2b,* based on their sequence similarity to each other and to *reggie-1* genes from other species (Table 2). We

named the newly found *reggie-1* from goldfish *reggie-1b* and renamed the previously cloned goldfish *reggie-1* (Schulte et al. 1997) as *reggie-1a.* Failure to amplify *reggie-1b* from the zebrafish could be due to an independent gene loss in the zebrafish genome or to technical difficulties such as inefficient primer annealing during PCR amplification.

New internal, zebrafish-specific primers were designed (Table 1, Fig. 2) and used on zebrafish cDNA in combination with some of the already mentioned degenerate primers, in order to extend the amount of coding sequence for the transcribed zebrafish genes. As a result, a total of 1188, 681, and 810 bp of coding sequence were obtained for *reggie-1a, reggie-2a,* and *reggie-2b,* respectively (Fig. 2). In addition, the *reggie-1* and *-2* introns of zebrafish and goldfish contribute 243 and 299 bp of DNA sequence, respectively.

The new zebrafish and goldfish sequences were aligned and compared to other *reggie* genes, in order to establish their evolutionary history. Percent aminoacid similarity between all known representatives of the *reggie-1* and *-2* groups are given in Table 2. For both groups of sequences, it can be observed that "*a*" or "*b*" genes are more similar between goldfish and zebrafish than to each other within either one species. For instance, *reggie-2a* genes of zebrafish and goldfish are 96.6% similar, while *reggie-2a* and *-2b* in zebrafish or goldfish are only 88.3

A.

```
                    ┌─────── Goldfish-reggie1a
              96/83 │
       100/93 ┌─────┤
          ┌───┤     └──────── Zebrafish-reggie1a
          │   │
          │   └─────────────── Goldfish-reggie1b
          │        ↖ loss of zebrafish reggie 1b?
          │
          │        ┌────────── Rat-reggie1
          │  100/100│
          ├─────────┤
          │         └───────── Human-Reggie1
          │
          └──────────────────── Fruitfly-reggie1
```

← (fish genome duplication)

B.

```
                    ┌─────── Zebrafish-reggie2a
             100/99 │
          ┌─────────┤
    92/83 │         └──────── Goldfish-reggie2a
     ┌────┤
     │    │         ┌──────── Zebrafish-reggie2b
     │    │  100/100 │
     │    └──────────┤
     │               └──────── Goldfish-reggie2b
     │
     │               ┌──────── Mouse-reggie2
     │       100/100 │
     │        ┌──────┤
     │ 100/100│      └──────── Rat-reggie2
     ├────────┤
     │        └─────────────── Human-Reggie2
     │
     └──────────────────────── Fruitfly-reggie2
```

─────── **0.05 changes**

Fig. 4. Distance (Kimura's two-parameter) and maximum parsimony consensus tree of *reggie-1* (**A**) and *reggie-2* (**B**) DNA sequences. Fruitfly, mouse, and human sequences have been renamed from their original "*flotillin*" designations to the "*reggie*" names used in this study (see Results). Trees were constructed using the neighbor-joining algorithm. The horizontal lengths of each branch are proportional to the estimated genetic distance (number of substitutions) between taxa. Bootstrap values (1000 replications) obtained by distance and parsimony methods are indicated on each branch, in that order and separated by a "/".

and 83.5% similar. Thus, the origin of "*a*" and "*b*" genes appears to predate the divergence of zebrafish and goldfish.

The Duplication of Fish reggie *Genes in Teleost Genomes*

The presence of additional *reggie* genes in zebrafish and goldfish, relative to tetrapods, could be explained by a gene/genome duplication in the common ancestor of these fish. Alternatively, the novel genes could have arisen through independent gene/genome duplication events in each of the two species. An ancestral duplication is more parsimonious than two independent duplication events in descending lineages. Phylogenetic analysis of all *reggie* genes was undertaken to decide between these two scenarios, based on the resulting tree topologies. If—as expected—the novel *reggie* copies arose in a teleost ancestor, homologous *reggie* sequences should cluster trans-specifically (Fig. 3A); if they arose

Table 2. (A) Percent similarity between known *Reggie-1* aminoacid sequences

	Drme-1	Mumu-1	Rano-1	Hosa-1	Dare-1a	Caau-1a	Caau-1b
Drme-1	—	59.7	61.3	61.3	63.6	62.2	64
Mumu-1	—	—	83	80	68	65	74
Rano-1	—	—	—	89	73	81	73
Hosa-1	—	—	—	—	74	81	73
Dare-1a	—	—	—	—	—	85	78
Caau-1a	—	—	—	—	—	—	88
Caau-1b	—	—	—	—	—	—	—

Species are designated by the first two letters of the genus and species name: Drme = Drosophila melanogaster, Rano = Rattus norvegicus, Hosa = Homo sapiens, Dare = Danio rerio, Caau = Carassius auratus.

(B) Percent similarity between known *Reggie-2* aminoacid sequences

	Drme-2	Mumu-2	Rano-2	Hosa-2	Dare-2a	Dare-2b	Caau-2a	Caau-2b
Drme-2	—	63.3	63.2	63.4	65.7	63.6	62	64.2
Mumu-2	—	—	99.5	97.9	88.7	78.3	79.2	79.5
Rano-2	—	—	—	97.7	88.7	78.3	79.4	79.5
Hosa-2	—	—	—	—	89.1	79	79.4	80.2
Dare-2a	—	—	—	—	—	83	96.1	84.3
Dare-2b	—	—	—	—	—	—	82.2	93.4
Caau-2a	—	—	—	—	—	—	—	81.9
Caau-2b	—	—	—	—	—	—	—	—

Species are designated by the first two letters of the genus and species name: Drme = Drosophila melanogaster, Mumu = Mus musculus; Rano = Rattus norvegicus, Hosa = Homo sapiens, Dare = Danio rerio, Caau = Carassius auratus.

independently in zebrafish and goldfish, then their sequences would group by species (Fig. 3B). Thus, the tree topology is indicative of the timing of the duplication. Distance and Maximum Parsimony methods using fruitfly sequences as outgroup, agree on the same tree topologies (Fig. 4A and B). Both *reggie-1* and *-2* trees are supported with high bootstrap values and clearly divide fish from mammalian sequences. Within the fish groups, sequences do not cluster according to species, but form instead clusters of "*a*" or "*b*" semiorthologous genes (Holland 1999) across species. Therefore, the existence of four *reggie* genes in zebrafish and goldfish is likely the result of a duplication event in a common ancestor of modern fish, after its separation from the lineage leading to tetrapods. Based only on our data, we cannot decide whether this event was a gene or an entire genome duplication. However, a whole-genome duplication in the common ancestor of all teleosts is strongly supported by data on *Hox* and other gene families (Amores et al. 1998; Postlethwait et al. 1998; Naruse et al. 2000; Málaga-Trillo and Meyer 2001; Málaga-Trillo et al. unpublished data). Although our data are obviously not sufficient to confirm the fish genome duplication hypothesis, they do add experimental evidence in its favor.

In the absence of mapping information, our results do not formally rule out the possibility that *reggie* genes may have duplicated as part of a partial duplication involving only Hox chromosomes (as opposed to a whole genome duplication), However, this can be demonstrated on the basis of genomic data from other organisms show-ing conserved mapping of *reggie* and *Hox* genes to different syntenic groups. In humans, *reggie-2* (*Flotillin-1*) maps to 6p2l.3, within the Major Histocompatibility Complex (*Mhc*) (MHC Sequencing Consortium 1999). The human *Mhc* is one of four extense paralogous regions that map to 6p21.3, 19p13.1, 9q33-q34, and 1q21-q25, independently of *Hox* clusters (chromosomes 2, 7, 12, and 17). The affiliation of *Reggie-2* within the *Mhc* is an ancient synteny conserved also in the zebrafish (Michalova et al. 2000). The ancestral linkage of *reggie-2* genes to the *Mhc* chromosomes, suggests that duplication events in vertebrate history involved more than just *Hox* chromosomes, and allows us to extend this assumption to the fish specific duplications. Human *Reggie-1* (*Flotillin-2*) does map to a *Hox* chromosome (17q12), but very distant from the *Hox B* cluster (17q21). This association is not representative of an ancestral genetic linkage between *Reggie* and *Hox* clusters, because in the fruitfly, neither *Reggie-1* nor *-2* (mapping to chromosomes 1 and 2, respectively) are linked to the single *Hox* complex (in chromosome 3). Thus, the presence of *Reggie-1* in human chromosome 17 is rather likely to be the consequence of an independent translocation event.

The Division Teleostei (bony fish) is made up of 38 Orders. One of these Orders—the Cypriniformes—includes the Cyprinidae, the largest family of freshwater fishes to which zebrafish and goldfish also belong (Nelson 1994). As these two genera represent only a fraction of the genetic and morphological diversity observed in modern teleosts, it will be necessary to identify ortholo-
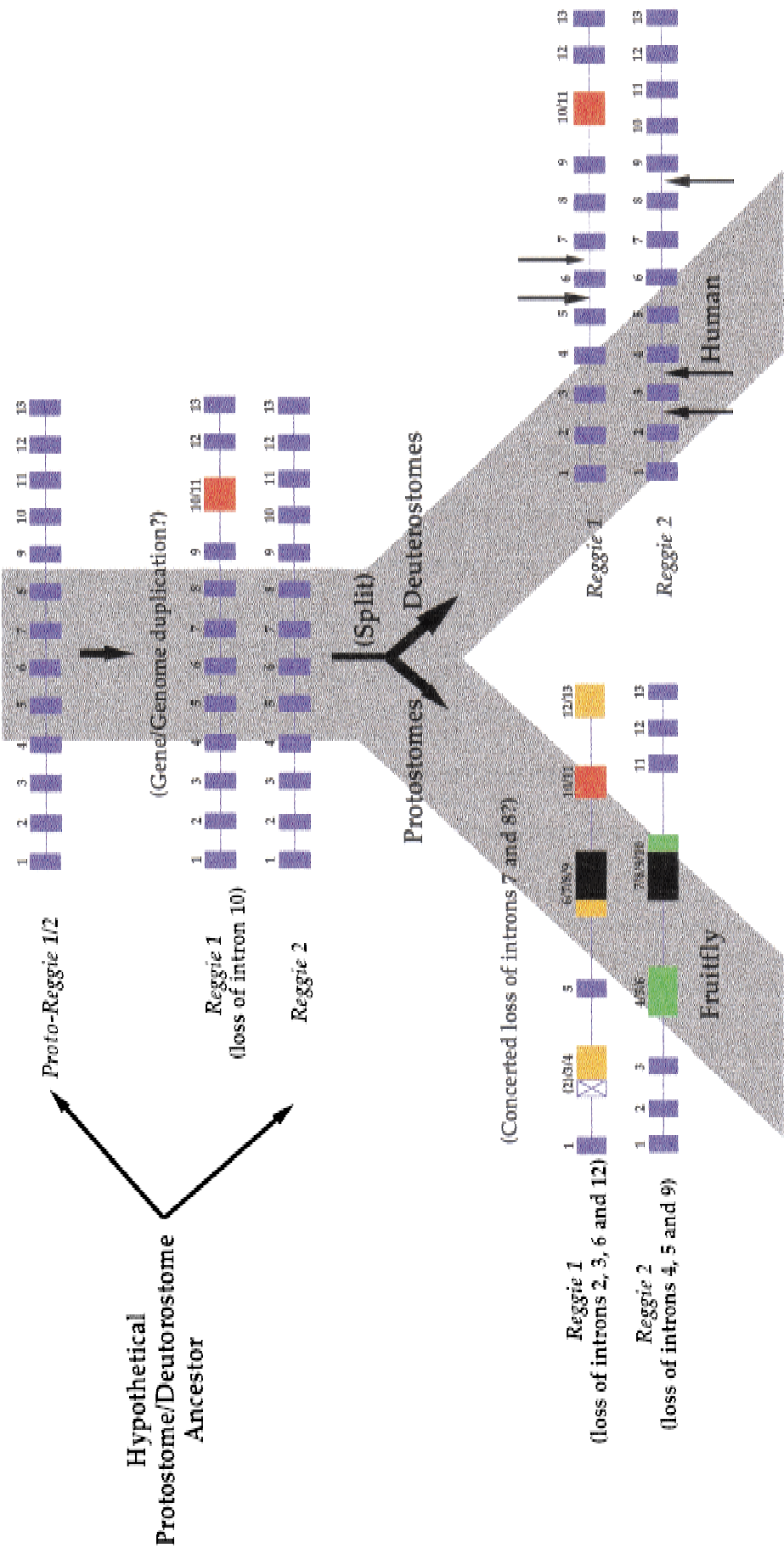
Fig. 5. Hypothetical scenario for the evolution of the intron-exon organization of *reggie* genes, based on the current configuration of human and fruitfly genes. The human and fruitfly gene structures were obtained by aligning the individual *reggie-1* and *-2* partial structures. Exons are represented by blue boxes; red boxes = intron losses shared by human and fruitfly genes; black boxes = concerted intron losses in fruitfly genes; orange boxes = fruitfly *reggie-1*-specific intron losses; green boxes = fruitfly *reggie-2*-specific intron losses. The fish introns identified in the present study are indicated by vertical arrows on the corresponding locations in the human genes. Exon 1 is non-conserved; it is only shown for illustrative purposes. Exon 2 is deleted in fruitfly *reggie-1* (box marked with a cross).

```
                 Exon 2                          Exon 3                                           Exon 4
               (                               (                                                (
Hosa Reggie 2    M---FFTCGPNEAMVVS g (Intron 2)   gt FCRSPPVM--VAGGRVFVLPCIQQIQ ag (Intron 3) )     g ISLNTLTLNVKSEKVYTRHGVP
Drme Reggie 2    MTWGFVTCGPNEALVVS g (Intron 2)   gg CCYMKPLL--VPGGRAFVWPVGQOVQ cg (Intron 3)       a ISLNTMTLQVESPCVVTSQGVP
Hosa Reggie 1    MG-NCHTVGPNEALVVS g (Intron 2)   gg GCCGSDYKQYVFGGWAWAWCISDTQ ag (Intron 3)        g ISLEIMTLQPRCEDVETAEGVA
Drme Reggie 1    .............NLS g ---------      gt GCCGSTKKRTIVGGWAWAWLVTDVQ cg ---------          a LSLNVMTLNPMCENVETSQGVP


                                                          Exon 5                                              (            )
                 )                                                                                        (  (Intron 5) )
Hosa Reggie 2    ISVTGIAQ (Intron 4)    VKIQGQNKEML-AAACQMFLGKTEAEIAHIALETLEGHQRAIMAHMTVE               (Intron 5) )      EIYKDRQK
Drme Reggie 2    ISVTGIAQ ---------     VKVQGQNEDMLL-TACEQFLGKSEAEINHIALVTLEGHQRAIMGSMTVE              ---------          EIYKDRKK
Hosa Reggie 1    LTVTGVAQ (Intron 4)    VKIMT--EKELLAVACEQFLGKNVQDIKNVVLQTLEGHLRSIL g (Intron 5) gg    TLTVEQIYQDRDQ
Drme Reggie 1    LTVTGVAQCKIMK (Intron 4)  --ADELLGTASEQFLGKSVKEIKQTILQTLEGHLRAIL g (Intron 5) ga       TLTVEEVYKDRDQ


                 Exon 6                                              Exon 7
               (                                                  (                                           )
Hosa Reggie 2    FSEQVFKVASSDLVNMGISVVSYTLKDIHDDQ (Intron 6) )   DYLHSLGKARTAQVQKDARIGEAEAKRDAGIR (Intron 7) )   EAKAKQEKVSA
Drme Reggie 2    FSKQVFEVASSDLANMGITVVSYTIKDLRDEE (Intron 6)     GYLRSLGMARTAEVKRDARIGEAEARAEAHIK ---------        EAIAEEQRMAA
Hosa Reggie 1    FAKLVREVAAPDVGRMGIEILSFTIK (Intron 6)           DVYDKVDYLSSLGKTQTAVVQRDADIGVAEAERDAGIR (Intron 7)  EAECKKEMLDV
Drme Reggie 1    FAALVREVAAPDVGRMGIEILSFTIK ---------            DVYDDVQYLASLGKAQTAVVKRDADAGVAEAANRDAGIR ---------   EAECEKSAMDV


                 Exon 8                                                         Exon 9
               (                                                             (
Hosa Reggie 2    QYLSEIEMAKAQRDYELKKAAYDIEVNTRRAQADLAYQLQ (Intron 8) )   VAKTKQQIEEQRVQVQVVERAQQVAVQEQEIARREKELEARVRKPAEAE
Drme Reggie 2    RFLNDTDIAKAQRDFELKKAAYDVEVQTKKAEAEMAYELQ ---------       AAKTKQRIKEEQMQVKVIERTQEIAVQEQEIMRRERELEATIRRPAEAE
Hosa Reggie 1    KFMADTKIADSKRAFELQKSAFSEEVNIK (Intron 8)                TAEAQLAYELQGAREQQKIRQEEIEIEVVQRKQIAVEAQEILRTDKELIATVRRPAEAE
Drme Reggie 1    KYSTDTKIEDNTRMYKLQKANFDQEINTA ---------                 KAQSQLAYELQAAKIRQRIRNEEIQIEVERRKQIEIESQEVQRKDRELTGTVKLPAEAE


                 Exon 10                                                    Exon 11
               (                                             (
Hosa Reggie 2    RYKLERLAEAE aa (Intron 9) g SQLIMQAEAEAASVR (Intron 10) )   MRGEAEAFAIGARARAEAEQMAKKAEAFQLYQEAAQLDMLLEK
Drme Reggie 2    KFRMEKLAEAN aa --------- g QRVVMEAEAEAES (Intron 10)          IRIRGEAEAFAIAAKAKAEAAEQMAMKAEAYREYREAAMVEMLLDT
Hosa Reggie 1    AHRIQQIAEGE aa (Intron 9) g VKQVLLAQAEAEKIR ---------          KIGEAEAAVIEAMGKAEAAERMKLKAEAYQKYGDAAKMALVLEA
Drme Reggie 1    AFRLQTLAQAK ca (Intron 9) a CQTIEGARAEAERIR ---------          KIGSAEAHAIELVGKAEAAERMRMKAHVYKQYGDAAIMNIVLES


                                   Exon 12                                                                    (Exon 13)
               )                 (                                                                          )
Hosa Reggie 2    LPQ (Intron 11)    VAEEISGPLTSANKITLVSSGSGTMGAAKVTGEVLDILTRLPESVERLTGVSISQ (Intron 12) )     VNHKPLRTA
Drme Reggie 2    LPK (Intron 11)    VAAEVAAPLSQAKKITMVSSGTGDIGAAKLTGEVLSIVNKVPELVKNITGVDIAR (Intron 12)       SVHAG
Hosa Reggie 1    LPQ (Intron 11)    IAAKIAAPLTKVDEIVVLSGDNS-----KVTSEVNRLLAELPASVHALTGVDLSK (Intron 12)       IPLIKKATGVQV
Drme Reggie 1    LPK (Intron 11)    IAAEVAAPLAKTDEIVLIGGNDN------ITNDVTRLVAQLPPSINALTGVDLSK                   V--LSKIPGAKA
```

Fig. 6. Aminoacid alignment of human and fruitfly *reggie-1* and *-2* coding sequences, including intron locations and phases. In those cases where intron phase is "1" or "2", the nucleotide sequence of the interrupted codons are shown flanking the intron. Dashes represent indels. The start codon of *reggie* genes is located in exon 2. Since exon 1 is noncoding and highly variable, the exon 1/intron 1 and intron 1/exon 2 boundaries were left out of our analysis. Exon 2 is deleted in fruitfly *reggie-1* (see Fig. 5).

gous *reggie* genes in other teleost families to confirm that the duplication of these genes occurred in a bony fish ancestor.

Evolution of the Intron-Exon Organization of reggie *Genes*

All goldfish and zebrafish sequences reported here were amplified from cDNA. In addition, goldfish *reggie-1b* and *-2b,* as well as zebrafish *reggie-2a* could be partially amplified from genomic DNA, uncovering the existence of at least five introns in the upstream half of these genes: A 98 and a 145 bp intron in *reggie-1;* and an 84, a 92, and a 123 bp in *reggie-2* (genomic sequences from the downstream half of fish *reggie* genes could not be amplified due to technical limitations). The degree of conservation in length, and location and phase between these fish introns and those of other organisms could not be immediately evaluated because the intron-exon structures of other *reggie* genes were not previously described. Nevertheless, recent data from genome sequencing projects (records unpublished but available at http://www.ncbi.nlm.nih.gov/Genomes/) provide extensive information on the intron-exon structure of human and fly *reggie* genes. We used this information to find out whether the partial gene structure of the fish genes obtained here corresponds with that of human and fruitfly genes, as well as to learn more about the mechanisms driving their intron evolution and gene structure.

We compared the cDNA and genomic sequences of human and fruitfly *reggie-1* and *-2* genes (see Materials and Methods for accession numbers) and inferred the size and location of introns relative to the reading frame. In contrast to their high degree of DNA sequence conservation, human and fruitfly *reggie* genes differ greatly in their intron-exon organization, fruitfly genes having fewer introns than their human counterparts (Fig. 5). Alignment of fruitfly and human coding regions shows general conservation of intron-exon boundaries and intron phases (Fig. 6). Except for two minor sliding events (at intron 4 and possibly at intron 10), the locations of all fruitfly introns correspond to those of the human ones. The use of human and fruitfly sequences for the comparison allowed us to identify shared (ancestral) intron positions and phases for nine out of the 11 introns analyzed. The difference in the number of introns between the two organisms is therefore likely to be the result of loss of introns in the lineage leading to fruitflies rather than the gain of introns in the lineage leading to humans. The latter scenario is non-parsimonious because it would require the repeated and independent de novo formation of human *reggie-1* and *-2* introns at exactly the same positions in eight out of 11 cases.

The fish introns reported here are generally shorter than their human fruitfly counterparts; they correspond to introns 5 and 6 in *reggie-1,* and to introns 2, 3 and 8

in *reggie-2* (Fig. 5, indicated by vertical arrows). In all cases, intron locations and phases are conserved between human and fish introns (Fig. 6). The presence of introns 6 and 8 in fish *reggie-1* and *-2,* respectively, is shared with human but not with fruitfly genes, supporting the notion that these introns are ancient and that they were lost in the lineage leading to invertebrates.

Using introns as cladistic characters, we attempted to reconstruct the evolutionary patterns of the intron loss in *reggie* genes, based on the intron-exon structure of human and fruitfly genes. We assume that the origin of *reggie-1* and *-2* predate the split between protostomes and deuterostomes (see phylogeny in Fig. 1), and that independent loss of introns is more parsimonious than their independent gain. *Reggie-1* and *-2* must have originated in a common ancestor of protostomes and deuterostomes, from the duplication of a *proto-reggie-1/2* gene with 13 exons; a subsequent loss of intron 10 occurred in the *reggie-2* gene of this hypothetical ancestor, as this loss is shared by human and fruitfly *reggie-2* genes. After the split between protostomes and deuterostomes, the lineage leading to the fruitfly underwent the concerted loss of introns 7, 8, and 9 in both *reggie-1* and *-2* genes, as well as the loss of introns 3, 6, and 12 in *reggie-1* and 4, 5, and 9 in *reggie-2* (Fig. 5). In addition, fruitfly *reggie-1* appears to have lost exon 2 (Fig. 6).

Much remains to be learned about the functional fate of the novel genes identified in this study. Our survey uncovered the presence of duplicated *reggie* transcripts in two- to three-day-old zebrafish embryos. We are currently examining their developmental expression patterns in zebrafish, and assessing their role in axon growth/regeneration and T-cell activation. A particular area of interest remains the question whether all four *reggie* genes of fish are expressed simultaneously in one microdomain, or rather in a combinatorial manner, as subsets that define distinct microdomains with different signaling properties, resulting in the development of novel regulatory interactions absent in mammals. The possible effects of these changes on axon growth/regeneration could hold much awaited answers about the developing nervous system of vertebrates and its inability to regenerate neurons at the adult state.

References

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish *hox* clusters and vertebrate genome evolution. Science 282:1711–1714

Bickel PE, Scherer PE, Schnitzer JE, Oh P, Lisanti MP, Lodish HF (1997) Flotillin and epidermal surface antigen define a new family

of caveolae-associated integral membrane proteins. J Biol Chem 272:13793–13802

Cho YJ, Chema D, Moskow JJ, Cho M, Schroeder WT, Overbeek P, Buchberg AM, Duvic M (1995) Epidermal surface antigen (MS17S1) is highly conserved between mouse and human. Genomics 27:251–258

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Galbiati F, Volonte D, Goltz JS, Steele Z, Sen J, Jurcsak J, Stein D, Stevens L, Lisanti MP (1998) Identification, sequence and developmental expression of invertebrate flotillins from *Drosophila melanogaster.* Gene 210:229–237

Gaze RM, Sharma SC (1970) Axial differences in the reinnervation of the goldfish optic tectum by regenerating optic nerve fibres. Exp Brain Res 10:171–181

The Genetics Computer Group I (1997) Wisconsin Package 9.1. Madison, WI

Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, van Eeden FJ, Jiang YJ, Heisenberg CP, Kelsh RN, Furutani-Seiki M, Vogelsang E, Beuchle D, Schach U, Fabian C, Nusslein-Volhard C (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio.* Development 123:1–36

Holland PW (1999) The effect of gene duplication on homology. Novartis Found Symp 222:226–236

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Lang DM, Lommel S, Jung M, Ankerhold R, Petrausch B, Laessing U, Wiechers MF, Plattner H, Stuermer CA (1998) Identification of reggie-1 and reggie-2 as plasmamembrane-associated proteins which cocluster with activated GPI-anchored cell adhesion molecules in non-caveolar micropatches in neurons. J Neurobiol 37:502–523

Málaga-Trillo E, Meyer A (2001) Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. Amer Zool, in press

Meyer A, Málaga-Trillo E (1999) Vertebrate genomics: more fishy tales about *Hox* genes. Curr Biol 9:R210–

Meyer A, Schartl M (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

Michalova V, Murray BW, Sultmann H, Klein J (2000) A contig map of the Mhc class I genomic region in the zebrafish reveals ancient synteny. J Immunol 164:5296–5305

Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, Shimada A, Wada H, Kusakabe T, Suzuki N, Kinoshita M, Kanamori A, Terado T, Kimura H, Nonaka M, Shima A (2000) A detailed linkage map of medaka, *Oryzias latipes.* Comparative genomics and genome evolution. Genetics 154:1773–1784

Nelson JS (1994) Fishes of the world, 3rd ed. J. Wiley, New York

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Berlin, New York

Ohno S (1999) The one-to-four rule and paralogues of sex-determining genes. Cell Mol Life Sci 55:824–830

Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar TS, Yelick P, Beier D, Joly JS, Larhammar D, Rosa F, et al. (1998) Vertebrate genome evolution and the zebrafish gene map. Nat Genet 18:345–349

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

Sambrook J, Maniatis T, Fritsch EF (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York

Schroeder WT, Stewart-Galetka S, Mandavilli S, Parry DA, Goldsmith L, Duvic M (1994) Cloning and characterization of a novel epidermal cell surface antigen (ESA). J Biol Chem 269:19983–19991

Schulte T, Paschke KA, Laessing U, Lottspeich F, Stuermer CA (1997) Reggie-1 and reggie-2, two cell surface proteins expressed by retinal ganglion cells during axon regeneration. Development 124:577–587

Stuermer CAO, Lang DM, Kirsch F, Deininger S, Wiechers MF, Plattner H (2001) GPI-anchored proteins and fyn kinase assemble in non-caveolar plasmamembrane microdomains defined by reggie-1 and reggie-2. Mol Biol Cell (in press)

Swofford DL (2000) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA

Volonte D, Galbiati F, Li S, Nishiyama K, Okamoto T, Lisanti MP (1999) Flotillins/cavatellins are differentially expressed in cells and tissues and form a hetero-oligomeric complex with caveolins in vivo. Characterization and epitope-mapping of a novel flotillin-1 monoclonal antibody probe. J Biol Chem 274:12702–12709

Wittbrodt J, Meyer A, Schartl M (1998) More genes in fish? Bioessays 20:511–515