Analysis of fish ZP1/ZPB homologous genes – evidence for both genome duplication and species-specific amplification models of evolution

S. J. Conner1,2 and D. C. Hughes3

1 Reproductive Biology and Genetics Group, Division of Medical Science, Medical School, University of Birmingham, Edgbaston, Birmingham B15 2TH, UK; 2 Assisted Conception Unit, Birmingham Women’s Hospital, Metchley Park Road, Edgbaston, Birmingham B15 2TG, UK; and 3 School of Environmental and Applied Sciences, University of Derby, Kedleston Road, Derby DE22 1GB, UK

The vertebrate egg envelope is composed of a family of related proteins, the zona pellucida (ZP) proteins, which are characterized by the presence of a conserved zona pellucida domain. Analysis of teleost fish ZP gene sequences has demonstrated that there are no direct orthologues of the mammalian ZPB and ZP1 genes, but that teleost fish contain multiple copies of two classes of genes (ZPXa and ZPXb) that are equally related to ZPB and ZP1. The two classes of genes are further distinguished by expression in liver or ovary, respectively, indicating there was probably an initial duplication event, followed by a switch to hepatic expression of one of the paralogues. This switch was followed in some species by additional amplification of one of the paralogues with the subsequent loss of the other. It is proposed that the expansion of the number of ZPXa and ZPXb genes and the acquisition of dual sites of synthesis are the result of an ancient polyploidization event, followed by additional species-specific gene amplifications.

Introduction

The generation of complexity by the process of gene duplication is widely accepted to occur (Ohno, 1970) and is believed to underlie the phenotypic diversity of vertebrates. Particular support for this hypothesis has been provided by analysis of the hox gene clusters. The presence of seven hox gene clusters in zebrafish has led to the suggestion that an additional genome duplication occurred in ray-finned fish (Actinopterygii) compared with other vertebrates (Amores et al., 1998). Recently, doubts have been raised about this model and the additional gene copies in teleost fish have been ascribed to frequent independent (species-specific) duplications (Robinson-Rechavi et al., 2001). Comparison of gene sequences across closely related species has demonstrated greater divergence in genes that are involved in sexual reproduction compared with those that are expressed in other tissues (Swanson and Vacquier, 2002), and vertebrate egg envelope genes in particular seem to be under positive Darwinian selection (Makalowski and Boguski, 1998; Swanson et al., 2001; Swanson and Vacquier, 2002).

Egg envelope genes provide an opportunity to assess models for evolution, as they are conserved throughout vertebrates. The zona pellucida proteins, named for the mammalian zona pellucida, share a common ZP domain, a large motif of approximately 260 residues with eight conserved cysteines. ZP domains are also found in a number of extracellular matrix proteins and are predicted to be involved in protein–protein interactions (Bork and Sander, 1992; Jovine et al., 2002). The prevailing view is that the mammalian zona pellucida is composed of three glycoproteins encoded by three genes ZP1, ZP2 and ZP3 (McLeskey et al., 1998), also known as ZPB, ZPA and ZPC, respectively (Harris et al., 1994). However, ZP1 and ZPB are not orthologous, and the human genome actually encodes four ZP genes: ZPB, ZP1, ZP2 and ZP3 (Hughes and Barratt, 1999). This contention is supported by the identification of distinct ZPB and ZP1 orthologues in chickens (Bausek et al., 2000). Additional ZP genes have also been identified in amphibians (Lindsay et al., 2001, 2002). ZPB/ZP1-related genes have been isolated from various fish species, but the use of varying, and in some cases incorrect, nomenclature has prevented easy categorization of these genes. In the present study, a phylogenetic analysis was performed on the available teleost fish ZPB/ZP1-related sequences. This analysis has revealed that the evolution of the ZP1/ZPB genes is more complicated than was previously realized and raises interesting questions about the composition, synthesis and assembly of the egg envelope in fish.
Table 1. Currently available fish ZPX protein sequences obtained from NCBI protein database

<table>
<thead>
<tr>
<th>Species</th>
<th>Annotation in database</th>
<th>Protein accession number</th>
<th>Site of expression</th>
<th>Predicted size (amino acids)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carp (Cyprinus carpio)</td>
<td>ZP2</td>
<td>CAA96575</td>
<td>Ovary</td>
<td>632</td>
<td>Chang et al., 1997</td>
</tr>
<tr>
<td></td>
<td>ZP2</td>
<td>CAA96574</td>
<td>Ovary</td>
<td>269</td>
<td>Chang et al., 1997</td>
</tr>
<tr>
<td></td>
<td>ZP2</td>
<td>CAA96573</td>
<td>Ovary</td>
<td>555</td>
<td>Chang et al., 1997</td>
</tr>
<tr>
<td></td>
<td>ZP2</td>
<td>CAA96572</td>
<td>Ovary</td>
<td>544</td>
<td>Chang et al., 1997</td>
</tr>
<tr>
<td>Goldfish (Carassius auratus)</td>
<td>ZP2</td>
<td>CAA96576</td>
<td>Ovary</td>
<td>580</td>
<td>Chang et al., 1997</td>
</tr>
<tr>
<td>Japanese medaka (Oryzias latipes)</td>
<td>ZPB-domain containing protein</td>
<td>AAD38905</td>
<td>Ovary</td>
<td>401</td>
<td>Kanamori, 2000</td>
</tr>
<tr>
<td></td>
<td>Choriogenin H</td>
<td>BAA13994</td>
<td>Liver</td>
<td>591</td>
<td>Murata et al., 1997</td>
</tr>
<tr>
<td></td>
<td>Choriogenin H minor</td>
<td>BAA76901</td>
<td>Liver</td>
<td>634</td>
<td>Sugiyama et al., 1998</td>
</tr>
<tr>
<td>Flounder (Pseudopleuronectes americanus)</td>
<td>ZP protein</td>
<td>AC59642</td>
<td>Liver</td>
<td>509</td>
<td>Lyons et al., 1993</td>
</tr>
<tr>
<td>Salmon (Salmo salar)</td>
<td>Eggshell protein</td>
<td>CAA04221</td>
<td>Liver</td>
<td>467</td>
<td>Oppen-Berntsen et al., 1990</td>
</tr>
<tr>
<td></td>
<td>Eggshell protein</td>
<td>CAA04220</td>
<td>Liver</td>
<td>439</td>
<td>Oppen-Berntsen et al., 1990</td>
</tr>
<tr>
<td>Trout (Oncorhynchus mykiss)</td>
<td>Vitelline envelope protein α</td>
<td>AAF71258</td>
<td>Liver</td>
<td>563</td>
<td>Hyllner et al., 2001</td>
</tr>
<tr>
<td></td>
<td>Vitelline envelope protein β</td>
<td>AAF71259</td>
<td>Liver</td>
<td>524</td>
<td>Hyllner et al., 2001</td>
</tr>
<tr>
<td></td>
<td>Zona radiata structural protein</td>
<td>AAK97529</td>
<td>Liver</td>
<td>403</td>
<td>Arukwe et al., 2001</td>
</tr>
<tr>
<td>Zebrafish (Danio rerio)</td>
<td>ZP2 (variant A)</td>
<td>AAK16578</td>
<td>Ovary</td>
<td>430</td>
<td>Mold et al., 2001</td>
</tr>
<tr>
<td></td>
<td>ZP2 (variant B)</td>
<td>AAK16577</td>
<td>Ovary</td>
<td>431</td>
<td>Mold et al., 2001</td>
</tr>
<tr>
<td></td>
<td>ZP2 (variant C)</td>
<td>AF331967</td>
<td>Ovary</td>
<td>432</td>
<td>Mold et al., 2001</td>
</tr>
<tr>
<td></td>
<td>ZP2</td>
<td>AAK16580</td>
<td>Ovary</td>
<td>430</td>
<td>Mold et al., 2001</td>
</tr>
<tr>
<td></td>
<td>ZP2</td>
<td>AAD49112</td>
<td>Ovary</td>
<td>428</td>
<td>Wang and Gong, 1999</td>
</tr>
<tr>
<td>Japanese eel (Anguilla japonica)</td>
<td>eSRS3</td>
<td>BAA36592</td>
<td>Ovary</td>
<td>446</td>
<td>Miura et al., 1998</td>
</tr>
</tbody>
</table>

Only complete sequences are shown.
ZP: zona pellucida.

Materials and Methods

Database analysis

Nucleotide and protein sequences showing similarity to ZP1/ZPB were identified using the BLAST search programme (http://www.ncbi.nlm.nih.gov/blast/) to search the non-redundant database. Human ZP1 (Hughes and Barratt, 1999) and human ZPB (NM_021186) sequences were used as query sequences.

Representative sequences from each fish species were used to search the dbEST database.

Alignment and phylogenetic analysis

Complete nucleotide and protein sequences were aligned using the program ClustalW (http://www.ebi.ac.uk/clustalw/) with default settings.

Phylogenetic trees were constructed using ClustalW (Higgins et al., 1994) and PIE (http://www.hgmp.mrc.ac.uk). Bootstrap analysis was constructed with 1000 replicates. Trees were plotted using Treeview (Page, 1996).

Results

Searches of the non-redundant nucleotide and protein databases identified a large number of sequences from fish showing homology to ZPB and ZP1 (Table 1), which were restricted to teleost fish. With the exception of the Japanese eel (Anguilla japonica), the fish belong to the Euteleosti subdivision. Partial sequences were not included in subsequent analyses. Initial investigations demonstrated that teleost fish do not have distinct ZP1 and ZPB proteins, but instead have ZP proteins that show almost equal homology to both the ZP1 and ZPB proteins of higher vertebrates. For example, the carp (Cyprinus carpio) protein (accession number CAA96575) shows 36% identity to human ZPI and 40% identity to human...
ZPB. The fish ZPB/ZP1-related genes will be referred to as ZPX. It is worth noting that a number of these genes had been incorrectly identified as ZP2-related, and therefore annotated as ZP2 (carp: Chang et al., 1997; zebrafish: Wang and Gong, 1999; Mold et al., 2001), even when the authors had noted greater similarity to the mammalian ZP1/ZPB genes (Mold et al., 2001). No orthologues of the mammalian ZP2 genes have been identified in fish to date.

In humans and other higher vertebrates, the ZP genes exist as single copy loci in the genome; in contrast, it is clear that there are multiple copies of the ZPX gene in a number of fish species (Table 1). BLAST searches of the dbEST database indicate that carp have at least six copies, the Japanese medaka (Oryzias latipes) has at least seven copies and the zebrafish (Danio rerio) has at least ten ZPX genes. As the data set is incomplete, it is not possible to predict how many ZPX genes any one fish species might have.

The multiple copies of carp, Japanese medaka (Oryzias latipes), salmon (Salmo salmar) and trout (Oncorhynchus mykiss) ZPX sequences show high levels of sequence identity at the DNA level except in exon 1. These differences can be attributed to the presence of a repeat domain of variable length. This motif comprises most of exon 1 and is rich in proline, glutamine and lysine residues. The number of repeats within the motif varies between the different ZPX genes even within a species. As a result, exon 1 varies considerably in length (354–967 bp). Repeats within a protein show greater similarity than between different proteins, and the repeat sequences differ more markedly between orders than between species. This sequence divergence is indicative of expansion of these repeats after gene duplication. The repeat domain is not present in all the ZPX genes identified. Of the three Japanese medaka sequences, AF128808 does not contain the repeat, and none of the zebrafish sequences do.

The five zebrafish (Danio rerio) ZPX sequences show an unusual pattern of sequence homology. Identity levels are extremely high (approximately 99% at nucleotide level) in exons 1–6, but there is significant sequence variation in exons 7 and 8. The lack of neutral sequence divergence in the first six exons of the zebrafish ZPX genes is curious and may represent homogenization as a result of gene conversion (Li, 1997).

A dendrogram produced from alignments of ZPX, ZP1 and ZPB protein sequences using a neighbour-joining algorithm is shown (Fig. 1). Trees produced using maximum likelihood analysis showed the same grouping of the sequences into distinct families (data not shown). Alignment of the corresponding DNA sequences gave analogous results (data not shown). The tree clearly shows the duplication and divergence of the ZP1 and ZPB genes from a common ancestor. The segregation of the teleost ZPX genes is reinforced by the observation that one group (ZPXa) is expressed in the ovary and the other (ZPXb) is expressed in the liver. Although representatives of both classes have been isolated from medaka, in many species only ZPXa (for example, zebrafish) or ZPXb (for example, salmon) homologues have been identified. A search of dbEST supported these findings, indicating that there is unlikely to be an ascertainment bias. It is worth noting that the only non-euteleost ZPX sequence available, from the Japanese eel, did not fall into either of the two groups and shows equal homology to both.

**Discussion**

Teleost fish sequences belonging to the ZPB/ZP1 class of genes were analysed and several important observations were made. It is apparent that there are multiple copies of these genes in the fish species that have been analysed, thereby confirming previous observations (Mold et al., 2001). These genes can be segregated into two groups (ZPXa and ZPXb) on the basis of sequence similarity, coincident with their site of expression (Table 1). The outlying position of the eel ZPX sequence is consistent with an ancient duplication after the divergence of the lineage leading to eels (Robinson-Rechavi et al., 2001). Analyses of other gene families such as Na+/K+-ATPase subunits support this model (Cutler and Cramb, 2001; Serluca et al., 2001). Teleost fish must have diverged from the vertebrate genealogy before the duplication that gave rise to separate ZPB and ZP1 proteins. The divergence of ZPB and ZP1 happened at some time before the separation of the Aves 150 million years ago.

After the ancient duplication in euteleosts, one of the duplicate copies appears to have substituted the ovary-specific expression common to most vertebrate egg envelope genes (McLeskey et al., 1998; Hughes and Barratt, 1999) with expression in the liver, under the control of oestrogen. The E box which directs expression to the ovary (Liang et al., 1997) was replaced with oestrogen response elements (Lyons et al., 1993). Independently, ZP3 is also expressed in the liver of a number of teleost fish species (Murata et al., 1995; Del Giacco et al., 1998; Hylner et al., 2001). Subsequently available sequence data indicate that different mechanisms may be operating in distinct euteleost species. Maintenance of copies of both genes ensured expression of ZPX in both liver and ovary (medaka), whereas loss of one of the genes led to a restriction of ZPX expression to either the liver (trout, salmon, winter flounder) or the oocyte (zebrafish, carp). There have also been recent gene amplifications within common ancestors of the fish species analysed, at least in part by tandem duplication (Mold et al., 2001).

Early models of gene duplication assumed that duplicated alleles are either quickly lost or that one of the copies rapidly acquires new, divergent function. These models have been superseded by the duplication–degeneration–complementation (DDC) model, which predicts that after gene duplication, each copy acquires...
mutations in its regulatory region resulting in expression of different members of the gene family in different tissues or at different stages of development. It is not clear what the advantage of expression of envelope genes in the liver might be, given that there is no evidence for any other functions of these proteins. There do not appear to be any obvious structural differences in the egg envelopes of different species of teleost fish (Dumont and Brummett, 1985). Multiple copies of a gene enable the production of larger amounts of protein, which is commonly seen in tumour progression or pesticide resistance in individual species (Guillemaud et al., 1996). There is certainly a requirement for large amounts of egg envelope protein. Fish oocytes are relatively large and many species spawn frequently during the reproductive season, each time producing large numbers of eggs; for example, a mature female medaka can lay up to 20 eggs every day for up to 2 months with the inner layer of each egg constituting approximately 25 µg protein (Yamagami et al., 1992). Gene amplification offers a simple way of responding to these physiological demands. An alternative possibility is that the availability of multiple copies of the ZPX genes allows for the generation of additional variation, which may increase the structural complexity and diversity of the egg coat. This is proposed to underlie the duplication of the carp fibroin-like substance and spider silk-forming protein genes (Guertet al., 1996; Chang and Huang, 2002). Like the ZPX genes, these genes also show variation in the number of copies of a repeat domain. In addition, the truncation of at least one of the carp ZPX sequences with a premature stop codon (Chang et al., 1997) indicates that ZPX genes are subject to a number of evolutionary processes. The acquisition of stop codons could provide additional functionality or may Fig. 1. Dendrogram based on ClustalW multiple sequence alignment of ZPX, ZPB and ZP1 protein sequences. Accession numbers in the NCBI database are given for each sequence. Bootstrap values expressed as percentages of 1000 replicates are given at each node. Families of zona pellucida (ZP) genes are labelled (ZPXa, ZPXb, ZP1 and ZPB). ZPXa proteins are expressed in the ovary, whereas ZPXb proteins are expressed in the liver.
be repaired by gene conversion events; a Flounder ZP3 cDNA sequence in GenBank also contains a premature stop codon (AJ310719).

Evidence for interactions between the egg envelope and spermatozoa in fish is sparse and it seems likely that in fish the role of the zona pellucida proteins is purelystructural. Teleost spermatozoa lack the acrosomal cap present in higher vertebrates, whereas the egg envelopes have an additional structure, the micropyle, which shows species-specific variation. As spermatozoa can enter the egg only through the micropyle, any sperm-binding molecules in the egg envelope would be expected to be found clustered around this opening. There is no suggestion of clustering of the zona pellucida proteins in fish egg envelopes around the micropyle. In the Japanese medaka, ZPX and ZP3 proteins are evenly distributed throughout the outer mucosal layer (Iwamatsu et al., 1997).

There is an additional important consequence of our observations about the teleost fish ZPX genes. Fish egg envelope proteins have been assessed as potential biomarkers for exposure to environmental oestrogens, as an alternative to vitellogenin. In several studies to date, expression of the egg envelope proteins has been detected at lower concentrations of oestradiol than vitellogenin, and expression has been induced more rapidly than vitellogenin in a number of fish species including salmon (Celius and Walther, 1998a), trout (Celius et al., 2000) and arctic char (Westerlund et al., 2001). Similar results have been obtained with the xeno-oestrogen α-zearalenol (Celius and Walther, 1998b). The pesticide o,p′-DDT (1,1,1-trichloro-2-[2-chlorophenyl]-2-[4-chlorophenyl]ethane) is also capable of inducing expression of egg envelope protein but not vitellogenin (Arukwe, 2001). It is worth noting that the effects of these agents are observed in not only female but also male fish, and that there are pathological consequences of exposure to oestrogen. Male fish, and female fish outside the reproductive cycle, accumulate fluid in the peritoneum containing significant amounts of the envelope proteins (Yamagami et al., 1992), which may be deleterious as the feminizing effects of exposure to xeno-oestrogens (Arukwe, 2001). We propose that the increased sensitivity of egg envelope genes as xeno-oestrogen biomarkers is not a result of increased activity of the regulatory regions, but instead reflects the increased copy number of these genes. In addition, it is likely that fish species that lack the liver-expressed paralogous group will not display this sensitivity to the effects of xeno-oestrogens.

Many of the difficulties in addressing the function of the vertebrate egg envelope genes have arisen from the use of confusing and inappropriate nomenclature (Hughes and Barratt, 1999), a situation which is magnified in the case of the fish egg envelope proteins. We suggest that any nomenclature should be dependent on the relationship between related genes and should always be supported by multiple alignment analysis. A recent study by Spargo and Hope (2003) analysed a cross-section of zona pellucida proteins. Although these workers did not focus on the fish genes and, thus, did not examine all the fish ZPX gene sequences, their phylogenetic tree clearly supports our findings despite using different analyses. It should be noted that Spargo and Hope (2003) used the term ‘ZPX’ in a different context to that used in the present study. As more fish gene sequences accumulate, clarification of the processes leading to increased complexity will be elucidated. Our analysis of the teleost fish ZPX gene sequences indicates that after the divergence of the lineage leading to eels, genome duplication in a euteleost ancestor probably created two paralogous classes of ZPX genes expressed in the liver and ovary, respectively. Subsequently, in different fish lineages there was either retention of both classes, or loss of one of the two classes, which has been followed by significant expansion of the copy number of the remaining gene or genes in recent ancestors of the species described herein. Therefore, these data support both the ancient genome duplication and the recent gene duplication models (Amores et al., 1998; Robinson-Rechavi et al., 2001). Confirmation of this model awaits the identification of these genes from more fish species. Elucidation of the functional significance of the expansion and varying sites of expression of the ZPX genes will prove an interesting challenge.

This work was supported by an MRC ROPA award and the NHS.

References


Arukwe A, Kullman SW and Hinton DE (2001) In vivo modulation of nonylphenol-induced zonagenesis and vitellogenesis by the antiestrogen, 3,3′,4,4′-tetrachlorobiphenyl (PCB-77) in juvenile fish Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology 129:1–10

Bausek N, Waclawek M, Schneider WJ and Wohlrab F (2000) The major chicken egg envelope protein ZP1 is different from ZPβ and is synthesized in the liver Journal of Biological Chemistry 275:866–872

Bork P and Sander C (1992) A large domain common to sperm receptors (Zp2 and Zp3) and TGF-beta type III receptor FEBS Letters 300:237–240

Celius T and Walther BT (1998a) Oogenesis in Atlantic salmon (Salmo salar L.) occurs by zonagenesis preceding vitellogenesis in vivo and in vitro Journal of Endocrinology 158:259–266


Downloaded from Bioscientifica.com at 02/28/2022 03:51:47AM via free access
Chang YS and Huang FL (2002) Fibroin-like substance is a major component of the outer layer of fertilization envelope via which carp egg adheres to the substratum Molecular Reproduction and Development 62 397–406


Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC and Sacco AG (1994) Cloning and characterization of zona pellucida genes and CNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families DNA Sequence 4 361–393


Liang L, Soyal SM and Dean J (1997) FIG-alpha, a germ cell specific transcription factor involved in the coordinate expression of the zona pellucida genes Development 124 4939–4947

Lindsay LL, Wallace MA and Hedrick JL (2001) A hatching enzyme substrate in the Xenopus laevis egg envelope is a high molecular weight ZPA homolog Development, Growth and Differentiation 43 305–313

Lindsay LL, Yang JC and Hedrick JL (2002) Identification and characterization of a unique Xenopus laevis egg envelope component, ZPD Development, Growth and Differentiation 44 205–212


Oppen-Berndsen DG, Helvik IV and Walther BT (1990) The major structural proteins of cod (Gadus morhua) eggshells and protein crosslinking during teleost egg hardening Developmental Biology 137 258–265

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers Computer Applications in the Biosciences 12 357–358

Robinson-Rechavi M, Marchand O, Escriva H and Laudet V (2001) An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes Current Biology 11 R458–R459


Received 26 February 2003.
First decision 10 April 2003.
Revised manuscript received 8 May 2003.
Accepted 19 May 2003.