Cloning of somatolactin α and β cDNAs in zebrafish and phylogenetic analysis of two distinct somatolactin subtypes in fish

**Yong Zhu*[1], John W Stiller*[1], Michael P Shaner*[1,2],
Angela Baldini[1], Jean-Luc Scemama[1]** and **Anthony A Capehart[1]**

[1]Department of Biology, East Carolina University, 1000 East Fifth Street, Greenville, North Carolina 27858, USA

[2]Mathematics and Science Department, Nash Community College, 522 N. Old Carriage Road, Rocky Mount, North Carolina 27804, USA

(Requests for offprints should be addressed to Y Zhu; Email: zhuy@mail.ecu.edu)

*(Y Zhu, J W Stiller and M P Shaner contributed equally to this work)

Abstract

Somatolactin (SL) is a pituitary hormone belonging to the growth hormone/prolactin superfamily, with recognizable homologues in all fish taxa examined to date. Although sequences from most fish share reasonably high sequence identity, several more highly divergent SLs have been reported. Goldfish SL and a second SL protein found in rainbow trout (rtSLP) are remarkably different from each other and also dissimilar to other SLs. It has been unclear whether rtSLP is a recent paralogue restricted to rainbow trout, or reflects a more ancient duplication of the SL gene, and whether it is related to the goldfish sequence. Here we report the cloning of two different zebrafish SL cDNAs, which share only 57·5% nucleotide and 47·7% deduced amino acid identities. One copy, designated zebrafish SLα (zfSLα), displays a typical range of sequence similarity to most other SLs. The other copy, zebrafish SLβ (zfSLβ), shows low identity to most other SLs; surprisingly, it is most similar to the divergent SL sequence from goldfish. The mRNAs of zfSLα and zfSLβ were expressed specifically in two distinct regions of the pars intermedia in zebrafish. Cells expressing zfSLα are located at the posterior pars intermedia, bordering the neurohypophysis, whereas zfSLβ is expressed in the anterior part of the pars intermedia, bordering the pars distalis. Phylogenetic analyses indicate that zfSLβ, goldfish SL and rtSLP all belong to the SL hormone family; however, along with the genes from eel and catfish, these divergent sequences form a group that is clearly distinct from all other SLs. These results suggest the presence of two distinct SL families, SLα and SLβ, which may trace back to a teleost genome duplication prior to divergence of the cyprinids and salmonids.

Journal of Endocrinology (2004) **182,** 509–518

Introduction

Somatolactin (SL) is a pituitary hormone that is a member of the prolactin (PRL)/growth hormone (GH) family (Ono *et al*. 1990). Most SLs isolated to date are reasonably similar in sequence (amino acid identity, 51–87%); the first clear exception reported was the SL from goldfish. Goldfish SL shares a peptide identity of only 35–48% with all other SLs (Cheng *et al*. 1997). Furthermore, goldfish SL is missing the third of seven conserved cysteine residues present in all other SLs known at that time. Its low similarity to other SLs, along with an absence of this third conserved Cys, suggested that the goldfish protein could represent a distinct form of the hormone. Unfortunately, in nearly all fish species examined, only one SL sequence had been reported, precluding the use of phylogenetic analyses that could have potentially discerned two SL subtypes. Recently, a rainbow trout SL-like protein (rtSLP) was identified (Yang & Chen 2003), in addition to the rainbow trout SL reported originally (Yang *et al*. 1997).

The rtSLP sequence is dissimilar to all known SLs (amino acid identities, 41–56%), including the sequence from goldfish. Thus, it is unclear whether rtSLP is a recent but rapidly evolving paralogue, unique to the trout lineage, or is an orthologue of the goldfish SL that traces back to more ancient gene duplication. Because a relationship between the rtSLP and goldfish SL is not immediately apparent, identification of a goldfish SL- and/or rtSLP-type orthologue in another species is required, along with a rigorous phylogenetic investigation of the SL hormone family.

Here we report the cloning of two highly divergent zebrafish SL (zfSL) cDNAs. One is a typical SL sequence, similar to those reported from most fish; the second, however, bears strong similarity to goldfish SL. Phylogenetic analyses place this second type of zfSL in a strongly supported clade with goldfish, eel and catfish SLs and rtSLP, separated from all other SLs. Based on their different patterns of expression and divergent phylogenetic histories, we designate these SL paralogues SL α-type (SLα) and SL β-type (SLβ).

Online version via http://www.endocrinology.org

Materials and Methods

Cloning of SL cDNAs

Thirty zebrafish, *Danio rerio*, were obtained from a local pet store and euthanized with MS-222. Whole brains including pituitary glands were removed immediately because of the small fish size. Total RNA was extracted by sonicating tissues in 1 ml TRIzol reagent (Gibco). 200 μl chloroform was added and the solution was vortexed vigorously. The aqueous layer was transferred to a clean RNase-free microcentrifuge tube, and total RNA was precipitated and redissolved in 100 μl water.

First-strand cDNA was synthesized from 1 μg total RNA using the GeneRacer kit (Invitrogen), in accordance with the manufacturer's instructions. Nested PCR was performed using a set of mixed primers (SLF1, 5′-GARAARCCTYCTVGAACRYC-3′; SLR1, 5′-GAC GRCAYTTVAGVAGYTTG-3′) designed from published SL sequences including goldfish. The first PCR was performed in 20 μl aliquots using a gradient Eppendorf Mastercycler. After a 2-min denaturation at 94 °C, the PCR cycle was repeated 30 times with a denaturation at 94 °C for 30 s, annealing at 45–55 °C for 30 s and elongation at 72 °C for 1 min. The products of the first PCR reaction were used as the templates for the second round of PCR. Using the nested primers (SLF2, 5′-GTCATCCA RCAYGCYGAGC-3′; SLR2, 5′-DRTCTCCATCTT RTGRGCRTC-3′), 20 μl aliquots were prepared. After a 2-min denaturation at 94 °C, the PCR cycle was repeated 30 times. The conditions for the PCR cycle were 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min.

Subcloning and DNA sequencing The amplified partial cDNA products were separated by agarose gel electrophoresis and ligated into a pGEM T-Easy vector (Promega). After the vector was transformed in XL1-Blue (Stratagene, La Jolla, CA, USA)–competent cells, 30 positive clones were selected by blue-white screening. Plasmid DNA was purified from bacterial cells using the QIAprep Spin Plasmid kit (Qiagen). All 30 plasmid DNAs were sequenced with forward and reverse universal primers using the Big-Dye Terminator kit and an ABI Prism 377 DNA sequencer (Perkin-Elmer, Wellesley, MA, USA).

5′ and 3′ rapid amplification of cDNA ends (RACE) Two distinct SL-like sequences were obtained from the 30 clones. Therefore, two sets of gene specific oligonucleotide primers were designed and synthesized from the two partial SL sequences. 5′ and 3′ RACE were performed using the GeneRacer kit as per manufacturer's directions. Nested PCR was performed and products were separated by agarose gel electrophoresis. The products were ligated into the TOPO TA vector (Invitrogen) and sequenced with forward and reverse vector-specific

primers. Sequence data were compiled using Sequence Navigator (ABI, Foster City, CA, USA). To obtain the full-length zebrafish cDNAs, two sets of gene-specific primers were designed for both SLs. The amplified products were gel-purified and ligated into a TA vector (pGEM-T easy vector). Multiple clones were selected and sequenced.

In situ *localization of two SL-producing cells*

Full-length zfSL cDNAs cloned into the pGEM-T easy vector were linearized with SpeI or SphI to generate templates for synthesizing sense or antisense probes respectively. Sense and antisense digoxigenin-labeled probes were generated using T7 or SP6 RNA polymerases by *in vitro* transcription according to the instructions of the manufacturer (Roche Applied Science, Indianapolis, IN, USA).

Zebrafish whole brains, including pituitaries, were removed and fixed immediately in 4% paraformaldehyde/ PBS overnight. The tissues were dehydrated, embedded in paraffin, cut into 5–8 μm-thick serial sections, and mounted on Superfrost slides (Fisher Scientific, Pittsburgh, PA, USA). Sections were processed, re-hydrated, post-fixed with 4% paraformaldehyde for 20 min, and then treated with 200 mM HCl for 10 min followed by 15 min digestion with 10 μg/ml protein kinase K. Hybridization was conducted with 10 ng antisense probe per slide in 60 μl buffer (50% deionized formamide/2 × SSC (where 1 × SSC is 0·15 M NaCl/0·015 M sodium citrate)/10% dextran sulfate/0·01% yeast RNA/0·02% SDS) overnight at 55 °C. Sections were washed in high-stringency buffer (1 × SSC/50% formamide) at 55 °C and 1 × SSC at room temperature. Sections were equilibrated with TBS (50 mM Tris, pH 7·5/150 mM NaCl) and blocked with 1 × Roche blocking solution containing 10% fetal bovine serum and 1% sheep serum in maleic buffer (100 mM maleic acid/150 mM NaCl, pH 7·5) for 1 h at room temperature. Sections were then incubated for 1 h at room temperature with anti-digoxigenin FAB fragment conjugated to alkaline phosphatase (Roche Applied Science; diluted 1/2000 with 1 × Roche blocking solution containing 10% fetal calf serum). Sections were washed with TBS, equilibrated in AP buffer and incubated with Nitro Blue Tetrazolium (NBT)/5-bromo-4-chloroindol-3-yl phosphate (BCIP) substrate (Roche Applied Science) overnight at 4 °C in a dark, humid chamber. Reactions were stopped by washing with TE buffer (10 mM Tris, pH 7·5/1 mM EDTA) twice, then fixed with 4% paraformaldehyde, washed twice with TE buffer, and mounted with 90% glycerol in PBS.

Phylogenetic analyses of two SL families

Phylogenetic analyses of the GH superfamily In order to verify the identity of both zebrafish sequences,

```
                                          GTGTGTGTTGTGTGT   15

GTGTTTGTAAAGCCGGCAAGTGACCATTGAGAAATCTGGGCAAAAATGAACACAGTTAAA    75
                                             M  N  T  V  K   -20
GTTCTGCAGGTGTGTGTGTGTGTGCTGATTCTGGGTCAGTTTTTTGGTTTCGGGTGCTGTT   135
 V  L  Q  V  C  V  C  V  L  I  L  G  Q  F  L  V  S  G  A  V     1
CCTCTGGACTGTAAGGATGACGCTGGATCTCGATGTGCTTCGATCTCTCAGGAAAAGCTT   195
 P  L  D [C] K  D  D  A  G  S  R [C] A  S  I  S  Q  E  K  L     21
CTGGATCGAGTGATCCAGCATGCTGAGCTCATTTACCGCGTCTCTGAGGAGTGCTGCACT   255
 L  D  R  V  I  Q  H  A  E  L  I  Y  R  V  S  E  E  C [C] T     41
TTGTTTGAGGATATGTTCGTCCCGTATCCACTGCATGTGTTGATCAACCAGGCTGGAAAC   315
 L  F  E  D  M  F  V  P  Y  P  L  H  V  L  I  N  Q  A  G  N     61
ACCTGCCACAGCAAACACATCCCCATCCCAACATCCAAGAGTGAGATTCAACAGATATCG   375
 T [C] H  S  K  H  I  P  I  P  T  S  K  S  E  I  Q  Q  I  S     81
GACTCATGGCTGCTTCACTCAGTGCTGTTCCTGGTTCAGTCGTGGATGGAGCCGCTGCTG   435
 D  S  W  L  L  H  S  V  L  F  L  V  Q  S  W  M  E  P  L  L     101
TATCTGCAGACGACTCTCGACCGATACGATGACGCTCCCAACGCTCTGCTCAGCAAAACC   495
 Y  L  Q  T  T  L  D  R  Y  D  D  A  P  N  A  L  L  S  K  T     121
AAGTGGGTTTCTGACAAACTGCTGAGCTTGGAGCAGGGGGTGGTGGTGCTCATCCGCAAG   555
 K  W  V  S  D  K  L  L  S  L  E  Q  G  V  V  V  L  I  R  K     141
ATGCTGGACGAGGGAATCATGCCATCCTCCACCATCTTCGATCACACTCAGAGCCCGTAT   615
 M  L  D  E  G  I  M  P  S  S  T  I  F  D  H  T  Q  S  P  Y     161
GACGGTCAGTTCCCGGAGGTGCTGGAGTCTGTGATCAGGGATTATCATCTGCTCACCTGC   675
 D  G  Q  F  P  E  V  L  E  S  V  I  R  D  Y  H  L  L  T [C]    181
TTCAAGAAAGACACGCACAAAATGGAGACCTTCCTCAAGCTGCTCAAGTGCAGACAGAGC   735
 F  K  K  D  T  H  K  M  E  T  F  L  K  L  L  K [C] R  Q  S     201
AACAAACTCAGCTGCCTGCCCCAATAAACACACACAGAGAGAG(A)26                804
 N  K  L  S [C] L  P  Q  *                                       209
```

Figure 1 Complete nucleotide sequence and deduced amino acid sequence of zfSLα. Numbers on the right indicate numbers of nucleotides and amino acids. The putative signal peptide and polyadenylation signal are underlined. *, Putative translation stop codon. Boxing indicates conserved cysteine residues.

phylogenetic analyses were carried out on an alignment of sequences from members of all three families (GH, PRL and SL). Three data sets were constructed comprising predicted amino acid sequences from 13 SLs, 13 GHs, and eight PRLs. Each was aligned using CLUSTAL X (Thompson *et al*. 1994), and alignments were adjusted to remove regions of ambiguity. Following optimization of individual group alignments, the three groups were re-aligned with emphasis on maintaining conserved core domains of the individual sub-alignments. This global alignment was used to produce a maximum-likelihood (ML) estimate of relationships among sequences. An initial analysis was carried out in TREEPUZZLE-50 (Strimmer & von Haeseler 1996) using a Jones, Taylor and Thornton (JTT) matrix for probabilities of change among amino acids, and an invariant + Γ model of rate variation among sites, with four discrete Γ rate categories. Likelihood estimates for amino acid frequencies and Γ parameters were then used to search for the tree with the ML using PROML in PHYLIP 3·6 (Felsenstein 1989), with 10 random sequence additions with global rearrangements.

To determine the strength of internal support for the tree topology recovered, and to further explore likelihood tree space, analyses were performed in MRBAYES 3·0 b4 (Huelsenbeck & Ronquist 2001) to determine posterior Bayesian probabilities of common nodes. Four simultaneous Markov chains were run, also under an invariant + Γ model (four categories) and a JTT substitution matrix.

Chains were begun with random *a priori* trees; subsequently trees were sampled from the posterior probability distribution every 10 generations. Stable likelihood values were achieved during the first 50 000 generations. An additional 350 000 generations were run and the first 200 000 were excluded from analysis of the posterior probability distribution. A majority-rule consensus tree then was constructed from the 20 000 sampled trees to determine Bayesian support values. In addition, 200 distance and parsimony bootstrap replicates were performed. Distance matrices were constructed in PHYLIP (PRODIST) using JTT weighting and used in neighbor-joining tree building. Parsimony bootstrap analyses were carried out using PROTPARS.

Phylogenetic analyses of SL families To investigate more rigorously the relationship of our zfSL sequences to other SLs, we performed further phylogenetic analyses of SL sequences in the absence of the two more distantly related hormone families. Zebrafish SLα and SLβ (zfSLα and zfSLβ) were aligned with 20 other reported SLs in CLUSTAL X, and sequences were adjusted manually to correct alignment errors and remove ambiguous regions. ML analysis was performed in TREEPUZZLE-50, using substitution parameters estimated as described above. Bayesian support values were calculated, also as described previously, except that chains were run for 800 000 generations and sampled

```
GTCAGTCAAACAGGACCTTCTTTTGCGCCGCAATGATGAAGAAAGCTACAGCTCTACAGG    60
                                 M  M  K  K  A  T  A  L  Q  V   -15
TTTGTGTGGTGGTTCTGGTCTGCTTAAAGAAGGCTGTGATCGGATCTCCAGTGGAGTGTC   120
 C  V  V  V  L  V  C  L  K  K  A  V  I  G  S  P  V  E [C] P    6
CAGACCAAGAGATTGGAGGGACGTCCTGCACAATCTCAGCGGCGAGAAGCTTCTGGACCGAG  180
 D  Q  E  I  G  G  T  S [C] T  I  S  A  E  K  L  L  D  R  A   26
CCGTTCAACATGCTGAGCTCATTTACCGCTTCTCGGAGGAAGCCAAGTTGCTGTTTGACG   240
 V  Q  H  A  E  L  I  Y  R  F  S  E  E  A  K  L  L  F  D  E   46
AGTTGCTCGACTTATTCGGAGGTATGAATCAGTATATTCCTGGAGGGGCTGTGTGTGCTC   300
 L  L  D  L  F  G  G  M  N  Q  Y  I  P  G  G  A  V [C] A  P   66
CCAAAACGGTGCCGGTTCCTTTGTCTAAAAGCGAAATTCAACAGATTTCCGACCGATGGC   360
 K  T  V  P  V  P  L  S  K  S  E  I  Q  Q  I  S  D  R  W  L   86
TTCTGCACTCAGTCCTGATTCTGGTTCAGTTCTGGATTGAGCCATTGGTGAATCTACAGA   420
 L  H  S  V  L  I  L  V  Q  F  W  I  E  P  L  V  N  L  Q  K  106
AATCTCTGGAGAACTATAAAAGTGCCCCAATCGGACTTCTTAGCAGGAACCAATGGATAG   480
 S  L  E  N  Y  K  S  A  P  I  G  L  L  S  R  N  Q  W  I  A  126
CTAGCAAACTATCAAGCCTAGAGGAGGGGCTGTTGGTTCTTATTAGACAGATTTTGGGTG   540
 S  K  L  S  S  L  E  E  G  L  L  V  L  I  R  Q  I  L  G  E  146
AAGGAGGCTTGGTGTTGGGTCCTCCAGAAGATGTATCGGATAACACTCTCTCTGTCGATG   600
 G  G  L  V  L  G  P  P  E  D  V  S  D  N  T  L  S  V  D  A  166
CTTTCGAGACAGTGAGACGAGACTACAGTGTGATCTACTGCTTCCGAAAAGACGCACACA   660
 F  E  T  V  R  R  D  Y  S  V  I  Y [C] F  R  K  D  A  H  K  186
AGATTCAGACGTTTCTCAAACTCCTGAAGTGCCGCCAGGTTGACAGGGAAAACTGCTCCC   720
 I  Q  T  F  L  K  L  L  K [C] R  Q  V  D  R  E  N [C] S  L  206
TCTTCTAGCAGAAAAACAACCCCAAACATTCTGCCAAATTATTTTGAGCGATTCTATTCT   780
 F  *                                                         207
GCCAAAGTTGCTTTAGCATTTGCACATTTTAAATTAAAAC**AATAAA**AATGCCTGCTTGAA   840
ATGCATGCCTGATT(A)₁₇                                            872
```

Figure 2 Complete nucleotide sequence and deduced amino acid sequence of zfSLβ. Numbers on the right indicate numbers of nucleotides and amino acids. The putative signal peptide is underlined and the polyadenylation signal is bold and underlined. *, Putative translation stop codon. Boxing indicates conserved cysteine residues.

every 100 generations. The final 5000 trees were used to construct a majority-rule consensus tree to estimate posterior clade probabilities. Both neighbor-joining and parsimony bootstrap percentages were determined as above, except that 1000 bootstrap replicates were performed in each case.

Analyses of outgroup rooting To determine whether outgroup rooting of the SL tree was likely to be a 'long-branch' phylogenetic artifact, 100 random sequences were generated in MacClade 3·0 (Maddison & Maddison 1992) using ML-estimated average amino acid frequencies. Each sequence was used in place of the outgroup (lungfish SL) in maximum-parsimony analysis, with 20 random sequence additions per random outgroup in PAUP 4·0 b.10 (Swofford 1997). Each random sequence's point of attachment was scored for all of the equally most-parsimonious trees recovered in each analysis (Stiller & Hall 1997). We also performed Shimodaira–Hasegawa (Shimodaira & Hasegawa 1999) tests in PROML on alternative root locations for the SL tree.

Results

Cloning of two distinct SL cDNAs

zfSLα sequence A 411 bp partial cDNA was amplified from zebrafish whole-brain cDNAs using degenerate primers. Subsequently, primers designed for 5′ and 3′ RACE amplified 750 and 240 bp products respectively. The full-length zfSLα was amplified for verification of the complete cDNA sequence and was found to be 778 bp, excluding the poly(A) tail. The complete cDNA sequence and predicted amino acids are depicted in Fig. 1. The zfSLα cDNA contains a 699 bp open reading frame encoding 233 amino acids. The predicated mature peptide has 209 amino acids. One polyadenylation site, AATAAA, is located at positions 758–763. The 5′ untranslated region is 60 bp long, while the 3′ untranslated region is 18 bp excluding the poly(A) tail. The calculated molecular mass and isoelectric point of the mature peptide are 23 970 Da and 5·73 respectively. The zfSLα mature protein has seven conserved cysteine residues, consistent with other reported SLα sequences.

zfSLβ sequence Using zebrafish whole-brain cDNAs, a 439 bp product was amplified with degenerate primers by nested PCR. The sequence of cDNA fragment amplified was most similar to goldfish SL (amino acid identity, 69·3%). Primers designed for 5′ and 3′ RACE amplified 386 and 364 bp products, respectively. Another set of primers was designed to amplify and verify the full-length zfSLβ cDNA sequence. The full-length zfSLβ cDNA is 854 bp, excluding the poly(A) tail (Fig. 2). The 5′ untranslated region is 32 nucleotides, while the 3′ untranslated region has 129 nucleotides, not including

Table 1 Comparison of percentage amino acid sequence identities of ZFRSLα (AF376857) and zfSLβ (AY221126) to all known SLs including goldfish (U72940), rtSLP (Yang and Chen 2003), channel catfish (AF267991), European eel (U63884), rainbow trout (Yang *et al.* 1997), chum salmon (D10640), flounder (M33695), Atlantic halibut (L02117), yellow perch (AY332490), rabbitfish (AB026186), orange spotted grouper (AY169406), Atlantic cod (D10639), gilthead seabream (L49205), red drum (AF062520), sole (U06753), lumpfish (L02118), Congo puffer (AF253066), white sturgeon (AB017200) and African lungfish (AF062744). Amino acid sequence of *Fugu* was deduced from genomic sequences of a genomic database (http://fugu.hgmp.mrc.ac.uk/)

	Amino acid sequence identity (%)	
	zfSLα	zfSLβ
Species		
zfSLβ	47·7	
Goldfish	45	69·3
rtSLP	53·4	45·3
Channel catfish	50·4	47·1
European eel	53·8	42·4
zfSLα		47·7
Rainbow trout	69·4	50·2
Chum salmon	69·8	49·8
Flounder	67·2	47·4
Atlantic halibut	66	46·8
Yellow perch	65·5	47·7
Rabbitfish	63·4	47
Orange spotted grouper	63	46·5
Atlantic cod	62·6	46·6
Gilthead seabream	62·6	45·3
Red drum	62·6	46·6
Sole	62·3	48·9
Lumpfish	61·2	47·7
Fugu	58·3	41·8
Congo puffer	55·3	44·6
White sturgeon	65·4	48·1
African lungfish	57·7	43·8

the poly(A) tail. The complete cDNA sequence and predicted amino acid sequence are depicted in Fig. 2. The zfSLβ gene has a 693 bp open reading frame encoding 231 amino acids from the start codon (ATG), nucleotide position 33, to the termination codon (TAA), position 728. The predicted signal peptide is located at positions 1–24. One polyadenylation signal, AATAAA, is located at nucleotide position 821–826. The calculated isoelectric point is 5·16 and the calculated molecular mass of the mature SL peptide is 23 249 Da. zfSLβ has six out of the seven cysteines that are conserved with all other reported mature SL sequences; however, the third highly conserved cysteine residue, present in zfSLα and most other SL sequences, has been replaced by a lysine in zfSLβ.

Sequence comparison of two distinct SLs

zfSLα and zfSLβ share only 47·7% amino acid and 57·5% nucleotide identities (Table 1). Unannotated zfSLα and

zfSLβ sequences are found on two separate contigs in the *Danio* genomic database (http://trace.ensembl.org/perl/ssahaview?server=danio_rerio). These results indicate clearly that the two SLs are encoded by two different genes. zfSLα cDNA is more similar to most known SLs ranging in identity from 45% (goldfish) to 70% (chum salmon) SL. In contrast, zfSLβ shares the highest peptide identity with goldfish SL (69·3%) and is much less similar to all other known SLs (amino acid identities, 41–50%). zfSLα shares amino acid identities of 33·1 and 25·1% with zebrafish GH (accession no. AY286447) and PRL (NM_181437) respectively, while zfSLβ is 31·4 and 22·9% identical to zebrafish GH and PRL. Thus, both in terms of its overall sequence composition and its conserved cysteine residues, zfSLβ appears to be related to the previously reported divergent SL from goldfish, while zfSLα is similar to SLs reported from most fish taxa.

In situ localization of two SL–producing cells

The locations of zfSL-expressing cells were determined by *in situ* hybridization of SL transcripts in the pituitary cells of adult zebrafish. No signals were obtained in tissue sections hybridized with SL sense probes (Fig. 3B and D). Pituitary cells displaying a strong hybridization signal to the zfSL antisense sequences were found in two distinct regions of the pars intermedia of zebrafish (Figs 3 and 4). zfSLα expression was detected in the posterior part of the pars intermedia bordering the neural tissue that penetrates the pars intermedia, whereas zfSLβ is expressed in the anterior part of the pars intermedia bordering the pars distalis (Figs 3 and 4).

Phylogenetic analyses of two SL families

Sequence comparisons show that zfSLβ had higher sequence identities to other SLs than to zebrafish GH and PRL; nevertheless, similarity comparisons do not provide conclusive evidence that zfSLβ and goldfish SL are members of the SL family. Therefore, phylogenetic analyses were used to more definitively classify the zebrafish paralogues, and to rigorously investigate the evolutionary relationships of zfSLα and zfSLβ to other pituitary hormones.

Phylogenetic analyses of the GH superfamily

Each of the three individual families (GH, PRL and SL) was recovered as a monophyletic group with strong internal support, regardless of the method of phylogenetic inference used (Fig. 5). All phylogenetic analyses nested both zfSL paralogues, goldfish SL and rtSLP sequences within the SL family, clearly distinct from GH and PRL, with strong statistical support (Fig. 5). Thus, none of these
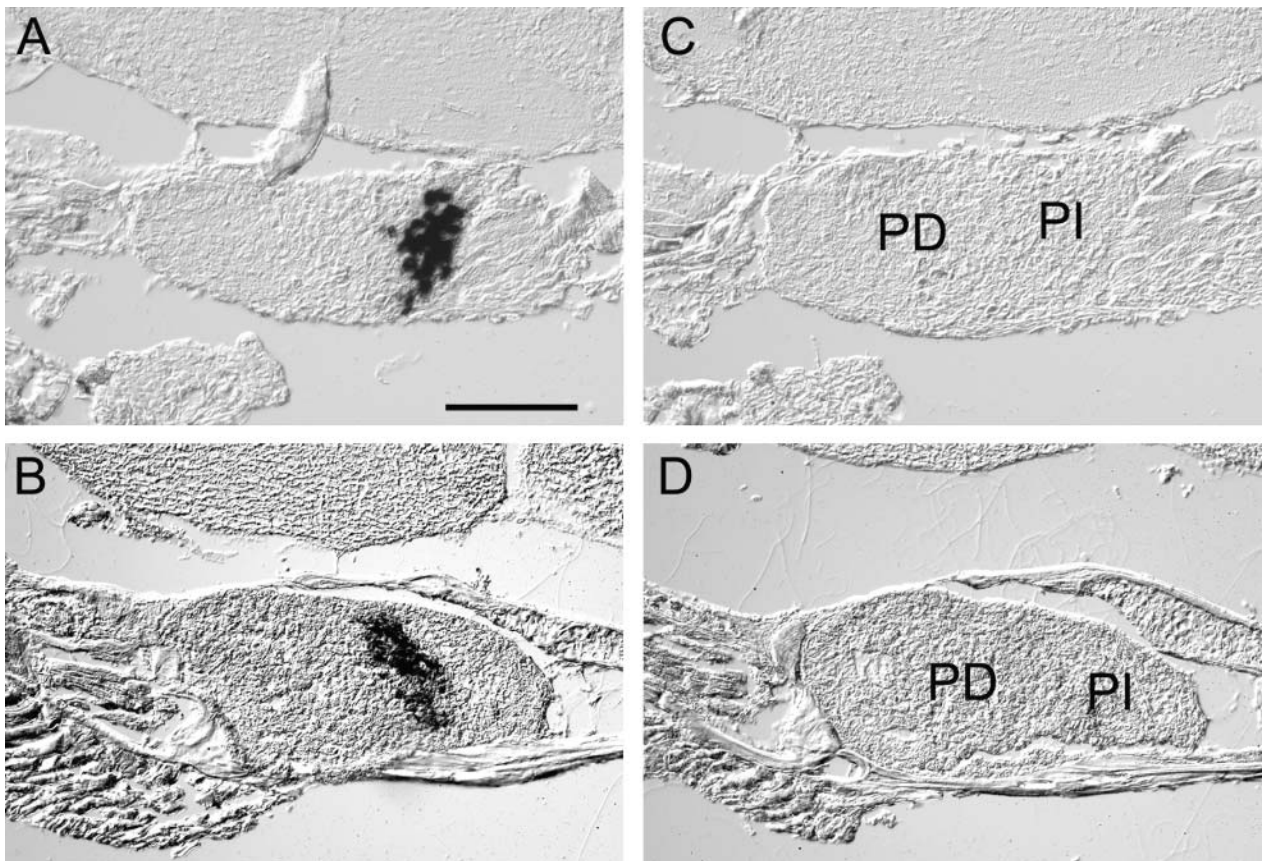
Figure 3 *In situ* hybridization of zebrafish pituitaries with zebrafish SL antisense and sense sequences. (A) and (C) show adjacent sections hybridized with digoxigenin-labeled zfSLα probes. Strong signals were obtained in the section (A) hybridized with zfSLα antisense sequence, whereas no signals were observed in the section (C) hybridized with zfSLα sense sequence. (B) and (D) shows adjacent sections hybridized with digoxigenin-labeled zfSLβ probes. Strong signals were obtained in the section (B) hybridized with zfSLβ antisense sequence, whereas no signals were observed in the section (D) hybridized with zfSLβ sense sequence. PI, pars intermedia; PD, pars distalis. Scale bar, 100 μm.

sequences is likely to represent a second copy of a hormone other than SL.

Phylogenetic analyses of SL families zfSLα branches with most other SL sequences in what appears to be reasonable expectation of the fish phylogeny. On the other hand, zfSLβ forms a second clade with goldfish, rtSLP and sequences from catfish and eel (Fig. 6). Both clades are recovered with strong statistical support, and with both the global as well as the SL-specific alignments (Figs 5 and 6). The presence of SL paralogues from both of these clades in zebrafish and rainbow trout, two distantly related species, suggests an ancient duplication within the teleost clade. Rooting fish SL evolution with an outgroup sequence from lungfish suggests that this duplication occurred after the divergence of sturgeon. Lungfish roots the SL tree on the branch leading to sturgeon in Bayesian, parsimony and distance bootstrap analyses, and sturgeon SL gives no

indication of 'long-branch' attraction in analyses of the random sequence behavior (Fig. 6).

Discussion

In the present study, we are the first to demonstrate the existence of two groups of SLs in teleosts that are produced by different pituitary cells. Since the discovery of SL (Ono *et al.* 1990, Rand-Weaver *et al.* 1991), no broad-scale phylogenetic analysis of the GH/PRL/SL families has been performed. Furthermore, goldfish SL has been excluded from previous phylogenetic analyses due to its low identity to other SLs (Zhu *et al.* 1999). The low amino acid identity of rtSLP, compared to other SLs, further complicates the classification of SLs in fish. Our analyses provide provocative evolutionary insights into the origins of the divergent SL copies in goldfish, rainbow trout and
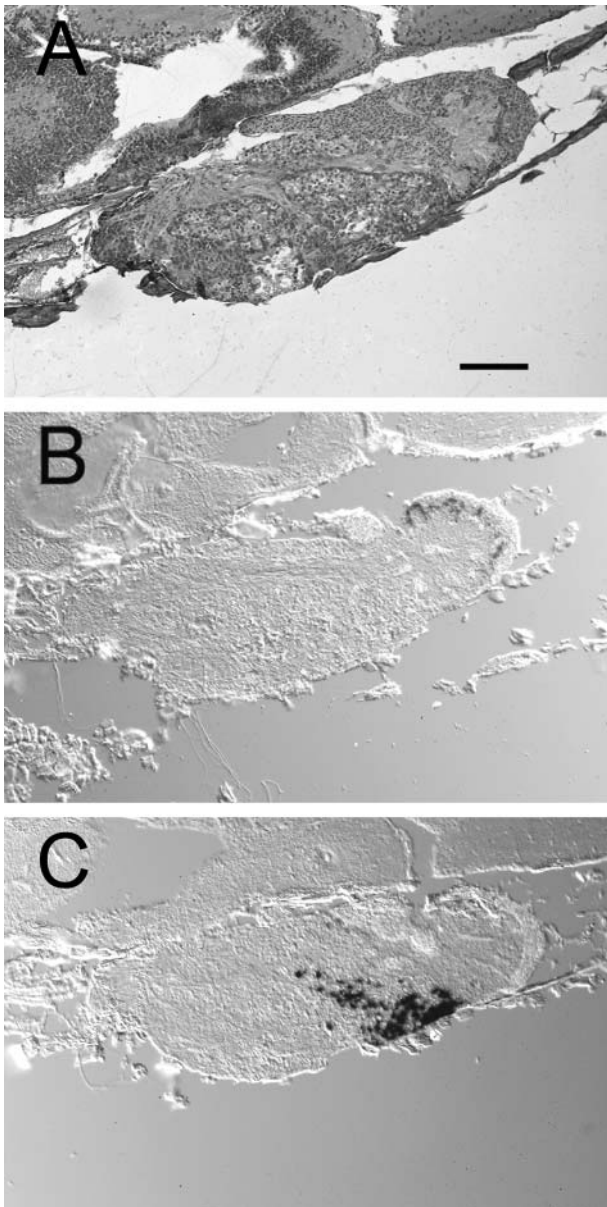
Figure 4 Localization of SLα- and SLβ-producing cells in adjacent sections of zebrafish pituitary. (A) Shown is a sagittal section with hematoxyolin/eosin staining. (B) and (C) show adjacent sections hybridized with a digoxigenin-labeled zfSLα (B) and zfSLβ (C) antisense sequences under Nomarski differential interference contrast. Scale bar, 100 μm.

zebrafish. We have identified two paralogous copies of zfSL, with distinct expression patterns. Furthermore, these paralogues fall into two distinct subgroups of the SL family, each of which is composed of diverse fish taxa. The single SL genes reported from goldfish, catfish and eel, along with zfSLβ and rtSLP, appear to belong to a newly classified SL group that we designate as the β group. The

other identified rainbow trout SL, along with zfSLα and all other SLs, belong to the α group.

Origin of SL paralogues

A fish-specific genome duplication was proposed after the discovery that fish had seven *Hox* gene clusters, almost twice as many as human and mouse (Amores *et al.* 1998, Prince 2002, Seoighe 2003). The hypothesis is becoming more widely accepted based on evidence from the recent release of the puffer fish and zebrafish genome sequences (van der Peer *et al.* 2003, Taylor *et al.* 2003, Venkatesh 2003). The strongly supported grouping of SL sequences from cyprinids, catfish, rainbow trout and eel, combined with the robust association of zfSLα with all other SL sequences, provides additional evidence for a genome duplication in the Actinopterygia (ray-finned fishes), with the subsequent maintenance of both paralogous SL copies in cyprinids and rainbow trout. Random sequence analyses of tree rooting suggest that the outgroup (lungfish) position between these two clades is not attributable to a simple 'long-branch' phylogenetic artifact (Fig. 6). Although some alternative rootings that fall between the rtSLP and zfSLβ cannot be rejected in Shimodaira–Hasegawa tests, any separation of the zfSLβ and goldfish sequences is rejected with significance. Therefore, there is significant support for classifying the zebrafish and goldfish SLβ genes as orthologues of each other, and as paralogues of the SLα clade. Nevertheless, it should be noted that, based on their empirical branch lengths, as well as the behavior of random sequences (Fig. 6), most members of the SLβ clade exhibit an accelerated rate of evolution. Further investigations are required to determine whether a clear duplication of SLs occurred in basal teleosts. In particular, identification of additional taxa containing both SLα and β paralogues is needed to more fully understand the age of the SL duplication and its ancestral distribution in fish.

Currently, it is unclear whether most fish possess both SL genes or have retained only one of the two paralogues. However, it is clear from complete genomic databases (http://fugu.hgmp.mrc.ac.uk/) that only SLα is present in *Fugu rubripes* and *Tetraodon nigroviridis*. A BLAST search of SLβ sequences only generates one single significant hit from each genome and that is the gene encoding SLα. The apparent loss of the SLβ gene in the puffer fish might indicate loss of the duplicated SLβ gene in one derived fish group. It is noteworthy that all fish that possess the SLβ gene live at least a significant portion of their lives in a fresh-water environment. Duplication of SL genes in those species may suggest important roles for SLs in an ancestral fish that invaded fresh water and required a second SL to help maintain homeostasis, including ion and water balance, body coloration and pigmentation (Fukamachi *et al.* 2004).
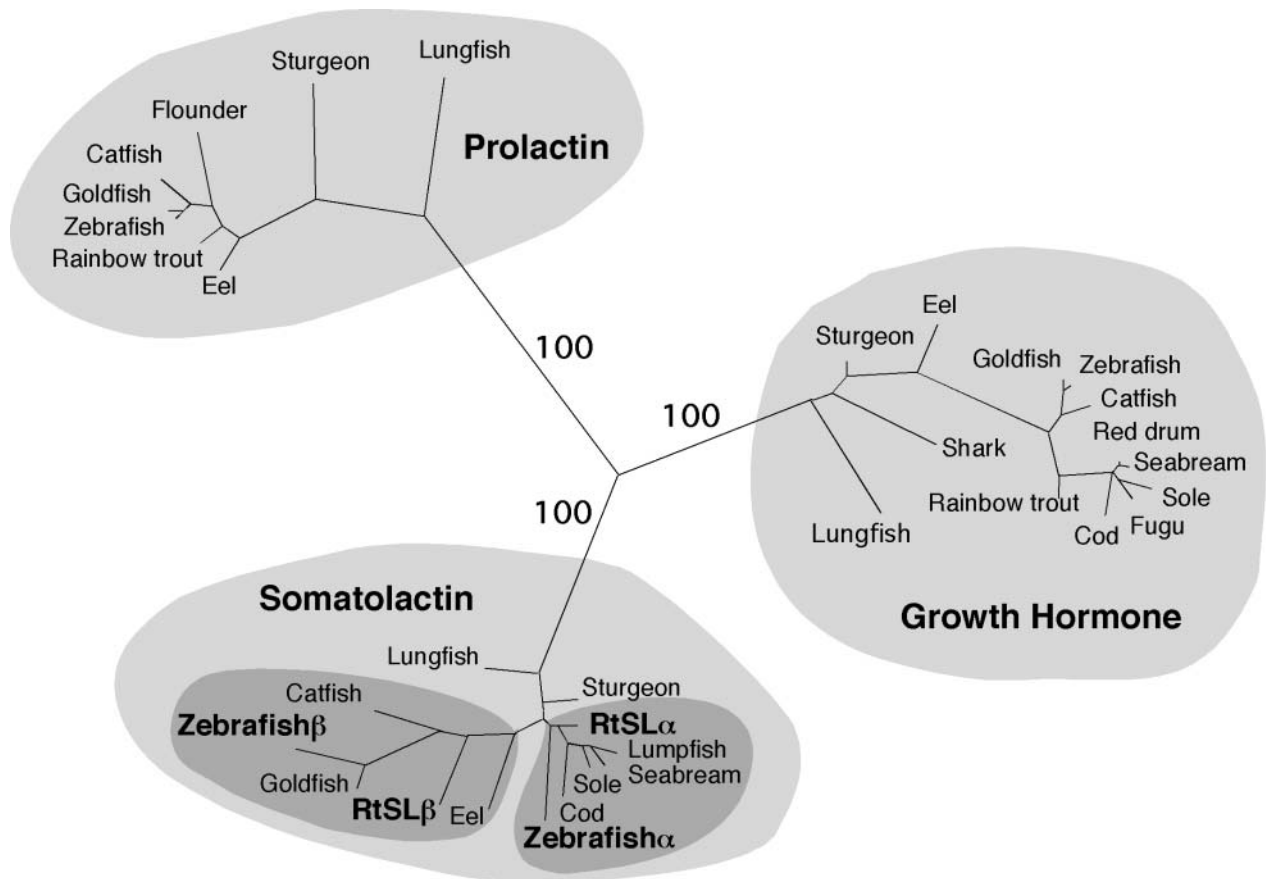
Figure 5 Maximum likelihood (ML) tree, with ML branch lengths, recovered from a global alignment of 34 hormones from the GH, PRL and SL families, using a JTT substitution matrix and an invariant + Γ model for variation among sites. A 100 next to the branches leading to each hormone family, and grouping of zfSLα and zfSLβ as specific orthologues of SL, indicate 100% support values in all analyses performed (Bayesian inference, parsimony bootstrap, and protein distance bootstrap).

The deduced amino acid sequence of zfSLα contains the seven conserved cysteines, a characteristic of most SLs. The zfSLβ, however, shares only six of these conserved cysteine residues; moreover, this third cysteine has been replaced by other amino acids in all members of the SLβ clade except for eel (indicated by an arrow in Fig. 6). Three disulfide bonds are formed among the first two and last four cysteines within the SL molecules, while the third cysteine is not involved in the formation of intra-peptide disulfide bonds (Rand-Weaver *et al.* 1991). The conservation of these six cysteines in both SLα and SLβ suggests that general structure and function are conserved between the two subtypes. The presence of the third cysteine in the SLα clade and its absence in β-type SLs, however, raises an intriguing question; that is, does loss of the third cysteine have important implications for SL structure and function? The rapid dimerization of SLα molecules may be mediated through formation of inter-molecular disulfide bonds between third cysteine residues (Zhu & Thomas 1995). Dimerization of hormone-receptor molecules is an import-

ant step in the activation of downstream signaling pathways; it may be that dimerization of the hormone molecules themselves also plays a role in regulating their activity. Further studies of the SL signal pathway, and of the effects of mutating the third cysteine residue, are clearly required.

Our preliminary results (Y Shaner & MP Zhu, unpublished observations) suggest that, like SLα in red drum (Zhu & Thomas 1996, 1998), SLβ is influenced by changes in illumination. This may indicate that SLα and SLβ share similar functions. However, low sequence identity of SLβ to α-type SLs, and differential expression in the pars intermedia, suggest that SLβ has acquired an alternative or additional set of functions. For example, SLα may be important during early development and SLβ may be important for maintaining homeostasis in adult fish (Yang & Chen 2003).

In summary, two types of SL were cloned from the same species, designated zfSLα and zfSLβ. Phylogenetic analyses suggest that these copies result from an ancient
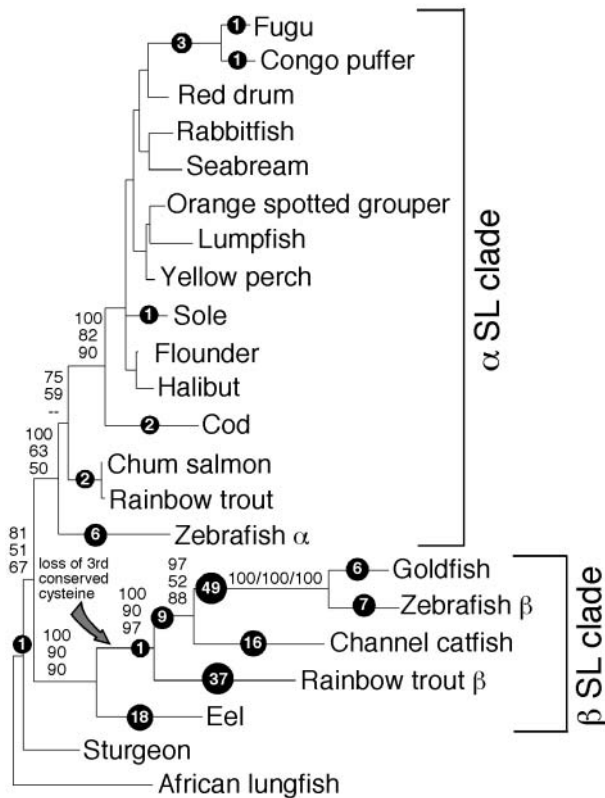
Figure 6 ML tree, with ML branch lengths, recovered from an alignment of 21 fish and one lungfish SL, using a JTT/invariant + Γ model of sequence evolution. Support values (Bayesian probability, parsimony bootstrap and protein distance bootstrap) relating to the relative positions of zfSLα and zfSLβ are shown above the relevant nodes. Black circles with white numerals are the number of times random sequences, used in place of the lungfish sequence, attached to that point on the tree in 100 separate parimony analyses using 100 different random sequences. All random sequences were constrained to contain average amino acid proportions for SL sequences calculated by ML.

duplication of SL genes in ray-finned fish. The duplication and maintenance of both SLs genes in some fish raise intriguing questions about the function and evolutionary history of these hormones, and suggest new avenues of investigation.

Acknowledgement

Funding

References

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince VE, Wang YL *et al.* 1988 Zebrafish hox clusters and vertebrate genome evolution. *Science* **282** 1711–1714.

Cheng KW, Chan YH, Chen YD, Yu KL & Chan KM 1997 Sequence of a cDNA clone encoding a novel somatolactin in goldfish, *Carassius auratus*. *Biochemical and Biophysical Research Communications* **232** 282–287.

Felsenstein J 1989 PHYLIP-phylogenetic inference package 32. *Cladistics* **5** 164–165.

Fukamachi S, Sugimoto M, Mitani H & Shima A 2004 Somatolactin selectively regulates proliferation and morphogenesis of neural-crest dervied pigment cells in medaka. *PNAS* **101** 10661–10666.

Huelsenbeck JP & Ronquist F 2001 MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17** 754–755.

Maddison WP & Maddison DR 1992 MacClade, ver 30. Sunderland, MA: Sinauer.

Ono M, Takayama Y, Rand-Weaver M, Sukata S, Yasunaga T, Noso T & Kawauchi H 1990 cDNA cloning of somatolactin, a pituitary protein related to growth hormone and prolactin. *PNAS* **87** 4330–4334.

van der Peer Y, Taylor JS & Meyer A 2003 Are all fish ancient polyploids? *Journal of Structural and Functional Genomics* **3** 65–73.

Prince VE 2002 The hox paradox: more complex(es) than imagined. *Developmental Biology* **249** 1–15.

Rand-Weaver M, Noso T, Muramoto K & Kawauchi H 1991 Isolation and characterization of somatolactin, a new protein related to growth hormone and prolactin from Atlantic cod (*Gadus morhua*) pituitary glands. *Biochemistry* **30** 1509–1515.

Seoighe C 2003 Turning the clock back on ancient genome duplication. *Current Opinion in Genetics and Development* **13** 1–8.

Shimodaira H & Hasegawa M 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16** 1114–1116.

Stiller JW & Hall BD 1999 Long-branch attraction and the rDNA model of early eukaryotic evolution. *Molecular Biology and Evolution* **16** 1270–1279.

Strimmer K & von Haeseler A 1996 Quartet puzzling: a quartet maximum–likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13** 964–969.

Swofford DL 1997 PAUP: Phylogenetic analysis using parsimony, ver 40b3a. Sunderland, MA: Sinauer.

Taylor JS, Brassch I, Frickey T, Meyer A & van de Peer Y 2003 Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Research* **13** 382–390.

Thompson JD, Higgins DG & Gibson TJ 1994 Clustal W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Research* **22** 4673–4680.

Venkatesh B 2003 Evolution and diversity of fish genomes. *Current Opinion in Genetics and Development* **13** 588–592.

Yang BY & Chen TT 2003 Identification of a new growth hormone family protein, somatolactin-like protein, in the rainbow trout (*Oncorhyncus mykiss*). *Endocrinology* **144** 850–857.

Yang BY, Arab M & Chen, TT 1997 Cloning and characterization of rainbow trout (*Oncorhynchuss mykiss*) somatolactin cDNA and its expression in pituitary glands and nonpituitary tissues. *General and Comparative Endocrinology* **106** 271–280.

Zhu Y & Thomas P 1995 Red drum somatolactin: development of a homologous radioimmunoassay and plasma levels after exposure to

stressors or various backgrounds. *General and Comparative Endocrinology* **99** 275–288.

Zhu Y & Thomas P 1996 Elevations of somatolactin in plasma and pituitaries and increased alpha-MSH cell activity in red drum exposed to black background and decreased illumination. *General and Comparative Endocrinology* **101** 21–31.

Zhu Y & Thomas P 1998 Effects of light on plasma somatolactin levels in red drum (*Sciaenops ocellatus*). *General and Comparative Endocrinology* **111** 76–82.

Zhu Y, Yoshiura Y, Kikuchi K, Aida K & Thomas P 1999 Cloning and phylogenetic relationship of red drum somatolactin cDNA and effects of light on pituitary somatolactin mRNA expression. *General and Comparative Endocrinology* **113** 69–79.