UNIVERSIDAD AUSTRAL DE CHILE
FACULTAD DE CIENCIAS

STUDY OF GENETIC AND EPIGENETIC DETERMINANTS OF GENE EXPRESSION REGULATION FOR BOTH SL GENES IN RESPONSE TO ESTROGEN IN COMMON CARP (*Cyprinus carpio*)

PhD Thesis

GUILLERMO EDUARDO VALENZUELA NIETO
VALDIVIA – CHILE

2018

STUDY OF GENETIC AND EPIGENETIC DETERMINANTS OF GENE EXPRESSION REGULATION FOR BOTH SL GENES IN RESPONSE TO ESTROGEN IN COMMON CARP (*Cyprinus carpio*)

Thesis submitted to the Faculty of Sciences of the Universidad Austral de Chile in partial satisfaction of the requirements for the Doctor of Philosophy in Sciences (PhD. in Sciences)

by

GUILLERMO EDUADO VALENZUELA NIETO
Valdivia - Chile

2018

UNIVERSIDAD AUSTRAL DE CHILE

FACULTAD DE CIENCIAS

INFORME DE APROBACIÓN DE TESIS DE DOCTORADO

La Comisión Evaluadora de Tesis comunica a la Dra. Leyla Cardenas, Directora de la Escuela de Graduados de la Facultad de Ciencias que la tesis de doctorado presentada por el candidato

GUILLERMO EDUARDO VALENZUELA NIETO

Ha sido aprobada en el examen de defensa de Tesis rendido el día 09 de octubre de 2018, como requisito para optar al grado de Doctor en Ciencias. Y, para que así conste para todos los efectos firman:

Profesor Patrocinante de Tesis:

Dra. Gudrun Kausel geb. Kamp ---------------------------------

Dr.Gernot Längst ---------------------------------

Comisión Evaluadora de Tesis:

Dr. Jaime Figueroa Valverde ----------------------------------

Dr. Alex Romero Zuñiga ----------------------------------

Dr. Alfredo Molina Sirguiado ----------------------------------

STUDY OF GENETIC AND EPIGENETIC DETERMINANTS OF GENE EXPRESSION REGULATION FOR BOTH SL GENES IN RESPONSE TO ESTROGEN IN COMMON CARP(*Cyprinus carpio*)

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT) DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

Vorgelegt von

GUILLERMO EDUARDO VALENZUELA NIETO

aus

VALDIVIA – CHILE

im Jahr
2018

Das Promotionsgesuch wurde eingereicht am: 20.08.18

Die Arbeit wurde angeleitet von:

Prof. Dr. Gudrun Kausel, Universidad Austral, Valdivia, Chile

Prof. Dr. Gernot Längst, Universität Regensburg

Unterschrift:

Die vorliegende Arbeit mit dem Titel

STUDY OF GENETIC AND EPIGENETIC DETERMINANTS OF GENE EXPRESSION REGULATION FOR BOTH SL GENES IN RESPONSE TO ESTROGEN IN COMMON CARP(*Cyprinus carpio*)

vorgelegt von Guillermo Eduardo Valenzuela Nieto

entstand unter der gemeinsamen Betreuung

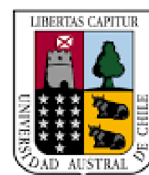
der Universität Regensburg

und

der Universidad Austral de Chile

im Rahmen des internationalen Promotionsprogramms *iP*UR

als Doppelpromotion



AGRADECIMIENTOS

A la Dra. Gudrun Kausel, por su incondicional y constante apoyo en todos los ámbitos, por su preocupación por mi desarrollo científico y personal, por ser un ejemplo de energía y buen corazón y por cultivar mi amor por la ciencia.

Al Dr. Marc Muller, por recibirme tanto en su casa como en su laboratorio y por todo lo que aprendí durante mi estadía en Bélgica.

Al Dr. Gernot Laengst, por recibirme en su laboratorio y en su empresa, por aceptar ser cotutor de esta tesis y por contribuir a mi formación científica.

Al Dr. Jaime Figueroa, por su apoyo y buen humor, por todo que aprendí de él sobre biología molecular y otros tópicos.

A los miembros del Laboratorio de Biología Molecular de Peces de hoy y ayer, en especial a Nano, Denise, Fran, CC, Nata, Dani, Gazu, Mati, Adolfo, Camo, Charly, Cata, Jamina, Rayi y Solomon.

A los muchachos de Regensburg, Elizabeth, Tom, Ingrid, Rodrigo, Hung y Cristian.

A con quienes compartí durante mi estadía en Liege, en especial al Dr. Maximiliano Figueroa y su equipo.

A con quienes colaboré científicamente durante la realización de esta tesis doctoral en materias ajenas a esta, en especial al Dr. Alejandro Rojas, a la Dra. Patricia Burgos, a la Dra. Julia Schwaiger y Gastón Vidal, por permitirme participar de sus líneas de investigación y de esa manera ampliar mi visión de la ciencia.

A Caroline por su hospitalidad, por su apoyo invaluable, por las cervezas compartidas y su amistad.

A Sandrita, por salvarme incontables veces, por hacernos la vida mas fácil a todos los de la escuela de graduados y por su bueno humor.

A los amigos de la montaña y en especial a mi cordada, Gabriel, Jano, Yertty y Christopher por su amistad y compañía.

A Hianara, por ser mi compañera de vida, en la ciencia, en la montaña y las buenas y malas, por su apoyo y cariño. Por las metas y sueños cumplidos y por cumplir.

A mi madrina Amelia por su apoyo a la distancia pero siempre presente y mis padres, Luis y Aida, por su cariño, dedicación y constante impulso para alcanzar la excelencia.

FUNDING

This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) Grant # 1070724, H2020 ITN722634 ProtectED, Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) PhD Fellowship # 21130511, International PhD Program at Universität Regensburg (iPUR), MECESUP AUS1203, DID UACh D #2015-02, Dirección de Postgrado Universidad Austral de Chile, and Escuela de Graduados, Facultad de Ciencias, Universidad Austral de Chile

TABLE OF CONTENTS

TABLE LIST	iii
FIGURE LIST	iv
ABBREVIATIONS	vi
1. ABSTRACT	viii
2. RESUMEN	x
3. INTRODUCTION	1
4. MATERIALS AND METHODS	21
4.1 MATERIALS	21
4.1.1 EQUIPMENT AND INSTRUMENTS	21
4.1.2 SOLUTIONS	22
4.1.3 SOFTWARE AND ONLINE TOOLS	24
4.1.4 BIOLOGICAL MATERIAL	25
4.1.4.1 ANIMALS	25
4.1.4.2 BACTERIAL STRAINS	25
4.2 METHODS	27
4.2 .1 ANIMAL TREATMENT	27
4.2 .2 MOLECULAR CLONING	27
4.2.2.1 ISOLATION OF PLASMID DNA	27
4.2.2.2 DETERMINATION OF NUCLEIC ACIDS QUANTITY AND PURITY	28
4.2.2.3 RESTRICTION ENZYME DIGEST	28
4.2.2.4 POLYMERASE CHAIN REACTION (PCR)	29
4.2.2.5 PURIFICATION OF PCR PRODUCTS	29
4.2.2.6 PURIFICATION OF DNA BY PHENOL/CHLOROFORM	30
4.2.2.7 INVERSE PCR	30
4.2.2.8 5`-RAPID AMPLIFICATION OF CDNA ENDS (RACE)	34
4.2.2.9 3`-RACE	37
4.2.2.10 GIBSON ASSEMBLY	37
4.2.2.11 AGAROSE GEL EXTRACTION	40
4.2.2.12 LIGATION	40
4.2.2.13 TRANSFORMATION	40
4.2.2.14 COLONY PCR	41

4.2.2.15 AGAROSE GEL ELECTROPHORESIS 42
4.2 .3 RNA EXTRACTION 42
4.2 .4 REVERSE TRANSCRIPTION 42
4.2 .5 RT-qPCR ANALYSES 42
4.2.6 ANTIBODY DEVELOPMENT 45
4.2 .7 DOT BLOT AND WESTERN BLOT 45
4.2 .8 IMMUNOHISTOCHEMISTRY 46
4.2 .9 PURIFICATION OF RECOMBINANT PIT-1 46
4.2 .10 DENATURING POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE) 47
4.2 .11 ELECTROMOBILITY SHIFT ASSAY (EMSA) 48
4.2.11.1 RADIOACTIVE EMSA 48
4.2.11.2 FLUORESCENT EMSA 49
4.2 .12 TRANSFECTION 52
4.2 .13 REPORTER ASSAY 52
4.2 .14 BISULFITE SEQUENCING 53
4.2 .15 CHROMATIN IMMUNO PRECIPITATION (CHIP) 55
4.2 .16 STATISTICAL ANALYSES 58

5. RESULTS 59
5.1. DIFFERENCES BETWEEN slα AND slβ GENES IN *Cyprinus carpio* 59
5.1.1 SL CODING SEQUENCES 59
5.1.2 SL EXPRESSION PATTERN 62
5.1.3 SL GENOMIC SEQUENCES 66
5.2 SL REGULATORY ELEMENTS 71
5.2.1 PIT-1 AND ER TRANSCRIPTION FACTORS BINDING SITES 76
5.2.2 INFLUENCE OF ER BINDING SITES ON ESTROGEN RESPONSIVENESS 80
5.3. METHYLATION PROFILE OF slα AND slβ IN RESPONSE TO ESTROGEN 82
5.4. slα AND slβ CHROMATIN STRUCTURE IN VIVO IN RESPONSE TO ESTROGEN 92
6. DISCUSION 100
7. CONCLUSION 119
8. SUPPLEMENTARY FIGURES 122
9. BIBLIOGRAFY 127

TABLE LIST

Table number	Name	Page
Table 1	Bacterial strain for cloning and protein expression	26
Table 2	Oligonucleotides to amplify promoter region of *sl* genes	33
Table 3	Oligonucleotide for amplification of cDNA ends	36
Table 4	Oligonucleotides for reporter assay constructions	39
Table 5	Oligonucleotides for qPCR quantification of both *sl* transcript levels	44
Table 6	Oligonucleotides for transcription factor binding assays	51
Table 7	Oligonucleotides for methylation assays	54
Table 8	Oligonucleotide used for quantification of ChIP DNA	57
Table 9	Transcription factor binding sites to both *sl* genes promoters	109

FIGURE LIST

Figure Number	Name	Page
Figure 1	Immunocytochemical localization of PRL, GH and SL in fish pituitary	3
Figure 2	Evolution of GH-PRL superfamily	6
Figure 3	Scheme of SL pathway	7
Figure 4	Hypothetical example of network evolution following a genome duplication	13
Figure 5	Estrogen effect on gene expression of hypophyseal and hypothalamic factors in male carp	19
Figure 6	*sl* transcript sequences	60
Figure 7	Amino acid sequence alignment of two SL carp genes	63
Figure 8	Differential spatial expression pattern of SL in carp pituitary	64
Figure 9	Differential expression of SL genes in adult carp	67
Figure 10	Identification of transcription start site of both *sl* genes	69
Figure 11	Genomic structure of *sl* genes	70
Figure 12	Carp *sl*α proximal promoter sequence	73
Figure 13	Carp *sl*β proximal promoter sequence	74
Figure 14	Divergences on *sl* genes regulatory sequences	75
Figure 15	Purification of recombinant PIT-1	77
Figure 16	*In vitro* binding of nuclear factors to predicted	79

	Pit-1 and ER binding sites in *sl*α and *sl*β promoters	
Figure 17	Pit-1 differential binding affinities to binding sites in *sl* promoter *in vitro*	81
Figure 18	*sl* reporter assay construct design	83
Figure 19	Analysis of *sl*β promoter activity using dual luciferase assay	85
Figure 20	Scheme for epigenetic characterization of *sl* response to estrogen	86
Figure 21	Methylation analyses procedure for *sl*α	88
Figure 22	Methylation analyses procedure for *sl*β	90
Figure 23	Methylation levels on *sl*α gene	93
Figure 24	Methylation levels of *sl*β gene	95
Figure 25	Primer design and standardization for qPCR of *sl* ChIP	98
Figure 26	Chromatin compaction state around *sl* promoters	99
Figure 27	Carp *sl*α and *sl*β proximal promoter scheme	110
Figure 28	Suggested *sl*β estrogenic response regulation	113
Supplementary Figure 1	Effect of BaP on pituitary *sl* gene expression.	122
Supplementary Figure 2	Epitope selection for carp *sl*α antibody	123
Supplementary Figure 3	gDNA fragmentation for immunoprecipitation.	125
Supplementary Figure 4	Methylation levels on both SL genes under basal conditions	126

ABREVIATIONS

Chromatin immunoprecipitation	ChIP
corticotropin-releasing factor	CRF
double stranded DNA	dsDNA
duplication-degeneration-complementation	DDC
Electromobility shift assay	EMSA
endocrine disrupters	EDs
Estrogen Receptor	ER
estrogen response element	ERE
gonadotropic hormones	GTH
gonadotropin-releasing hormone	GnRH
growth hormone	GH
Histone 3 trimethylated at Lysine 4	H3K4me3
Histone 3 trimethylated at Lysine 9	H3K9me3
melanotropin	MSH
Millions Years Ago	MYA
pars intermedia	PI
Polymerase chain reaction	PCR
Prolactin	PRL
proximal pars distalis	PPD
Quantitative reverse transcription PCR	RT-qPCR
Rapid Amplification of cDNA Ends	RACE
rostral pars distalis	RPD

Somatolactin	SL
thyroid-stimulating hormone	TSH
Untranslated Region	UTR

1. ABSTRACT

Somatolactin (SL), a fish hypophyseal hormone involved in background adaptation, reproduction and fatty acid metabolism might be affected by estrogenic endocrine disruptor compounds. Two *sl* transcripts were detected in pituitary of *Cyprinus carpio*, *sl*α and *sl*β, with only about 47,6% identity of coding sequences, but conserved key features in the derived amino acid sequences such as cysteine number and position. At basal conditions *sl*α showed higher mRNA levels than *sl*β and SLβ was immunodetected in different subregions of *pars intermedia* indicating spatiotemporal divergent expression patterns. When exposed to estrogen only *sl*β but not *sl*α responded with increased expression in pituitary of male adult carp to 17β-estrogen treatment respect to control as shown by RT-qPCR analyses. The in depth comparative analyses of regulatory elements of *sl*α and *sl*β revealed highly divergent, promoter regions besides the presence of Pit-1 binding sites in both, specially the one located in the neighborhood of TATA box was conserved. Indeed, in *sl*β but not in *sl*α promoter a cognate Estrogen Response Element (ERE) half site was found. A functional assay on GH3/BH6 cells demonstrated that this ERE half site was directly related with differential response of *sl* genes to 17β-estrogen. No mayor differences were detected on CpG methylation of promoter and coding sequences of both *sl* genes in response to the estrogen treatment. However, *sl*α showed a decreased methylation level respect to *sl*β promoter under basal conditions, suggesting a role of methylation on regulation of *sl* expression. *sl*β promoter showed a decrease in inactive chromatin marks in response to estrogen that correlated with increased transcript levels. Taken together these data suggest that *sl* paralogs diverged at genetic and epigenetic levels since last duplication event.

Keywords: Somatolactin, Duplicated genes, Estrogen response, endocrine disruptors, *Cyprinus carpio*.

2. RESUMEN

SL, una hormona hipofisaria de peces implicada en la adaptación al color del fondo, reproducción y el metabolismo de los ácidos grasos se ha descrito que puede verse afectada por los compuestos disruptores endocrinos de naturaleza estrogénica. Se detectaron dos transcritos *sl*, en la hipófisis de *Cyprinus carpio*, *sla* y *slβ* con un 47,6% de identidad en sus regiones codificantes, sin embargo, se conservan características clave como el número y la posición de los residuos cisteína. En condiciones basales *sla* mostró mayores niveles de mRNA que *slβ*, Además SLβ fue inmundetectado in diferentes subregiones de *pars intermedia* indicando un patrón de expresión espacial divergente. Por otra parte, solo *sl*β pero no *sl*α, respondió con un aumento en su expresión en hipófisis de carpa adulta macho al tratamiento con 17β-estrógeno respecto al control, como se mostró mediante análisis RT-qPCR. El análisis comparativo en profundidad de los elementos regulatorios de *sl*α y *sl*β reveló regiones promotoras altamente divergentes, a excepción de la presencia de sitios de unión de Pit-1 conservados en ambos, especialmente uno localizado en la vecindad de la caja TATA. Se encontró además un medio sitio del elemento de respuesta de estrógenos (ERE) en el promotor de *sl*β pero no en el promotor *sl*α. Un ensayo funcional en células GH3/BH6 demostró que este medio sitio ERE estaba directamente relacionado con la respuesta diferencial de los genes de *sl* a 17β-estrógeno. No se encontraron mayores diferencias en la metilación de CpGs en los promotores ni en la región codificante de ambos genes *sl* en respuesta al estrógeno. Sin embargo, *sl*α mostró bajos niveles de metilación respecto al promotor *sl*β en condiciones basales. Lo que sugiere que la metilación juega un papel en la regulación de la expresión *sl*. El promotor *sl*β mostró una disminución en los niveles de marcador de cromatina inactiva en respuesta al estrógeno que se correlaciona con niveles

de transcripción aumentados. Tomados en conjunto, estos datos sugieren que los parálogos de *sl* divergieron a niveles genéticos y epigenéticos desde el último evento de duplicación.

Palabras Clave: Somatolactina, Estrógeno, Genes Duplicados, Disruptores endocrinos, Carpa, *Cyprinus carpio.*

3. INTRODUCTION

Fish are subject to numerous and sometimes drastic changes in the conditions of the environment they inhabit, so the homeostasis of fish must be regulated closely by the neuroendocrine axis, through the release of hormones from the master gland, the pituitary gland (Kausel *et al.,* 1999; Figueroa *et al.,* 2005). The pituitary gland functions as a relay between the hypothalamus and peripheral target organs that regulate basic physiological functions, including growth, the stress response, reproduction, metabolism and lactation (Zhu *et al.,* 2004).

The endocrine axis involves the release of the hormone from an endocrine gland into the circulation, in response to an external stimulus. The hormones reach the cells of the body but exert changes only in the target organs that express the receptors of these hormones and manage to transduce the signal into the cell. There are at least 3 levels of regulation that govern the normal functioning of the endocrine axis: The appropriate synthesis and release of the hormone; the maintenance of hormone levels in the circulation and the expression of appropriate levels of functional receptors in the target organ. Loops of positive and negative regulation operate to accentuate or mitigate, respectively, the action of the hormone through the regulation of its release or the sensitivity of the organ in response to it. The malfunction or disregulation of any of these three levels of control or the disruption of some key regulatory loop, could result in the initiation or progression of an endocrine disease (Kronenberg and Williams, 2008)

The hypophysis, the pituitary master gland, is divided into adenohypophysis, originating from ectodermal germ layer by invagination of the roof of the developing mouth forming Rathke's pouch, and neurohypophysis, the posterior lobe which originates from neural ectoderm (Rosenfeld *et al.,* 2000;

Zhu *et al.,* 2007). In teleost cell lines producing different hormones are produced in distint areas, the adenohypophysis is subdivided into two zones: *rostral pars distalis* (RPD), where the lactotrope cells producing Prolactin (PRL) and corticotropes cells producing ACTH are located; *proximal pars distalis* (PPD) (Fig.1), where the somatotropes cells, producing growth hormone (GH), gonadotropes cells, producing gonadotropic hormones (GTH) and thyrotropes cells, producing thyroid-stimulating hormone (TSH) are located (Pogoda and Hammerschmidt, 2007; Zhu *et al.,* 2004). In neurointermediate lobe, *pars intermedia* (PI) melanotropes cells, producing melanotropin (MSH) and the cells producing Somatolactin (SL) are located (Kaneko, 1996). The embryonic development of the gland and the regulation of the expression of the hormones produced GH, PRL, TSH-beta subunit and in fish additionally SL is under the control of a specific transcription factor of the pituitary, Pit-1, which in turn regulates its own expression, which is modulated by environmental changes (Kausel *et al.,* 1999; Dasen and Rosenfeld, 2001).

Because of the absolute requirement of Pit-1 for somatotroph, lactotroph, and thyrotroph cell lines development and specific gene expression, inactivating mutations of the gene result in a spectrum of pituitary hormone deficiencies (Pfäffle *et al.,* 1992). Pit-1 belongs to a group of transcription factors that have the ability to overcome chromatin restriction, at least on nucleosome level, the master regulators (also called 'pioneering factors'), whereas a second tier of transcription factors is thought to primarily gain access to binding sites that are already 'made accessible' by master regulators (Pham *et al.,* 2013).

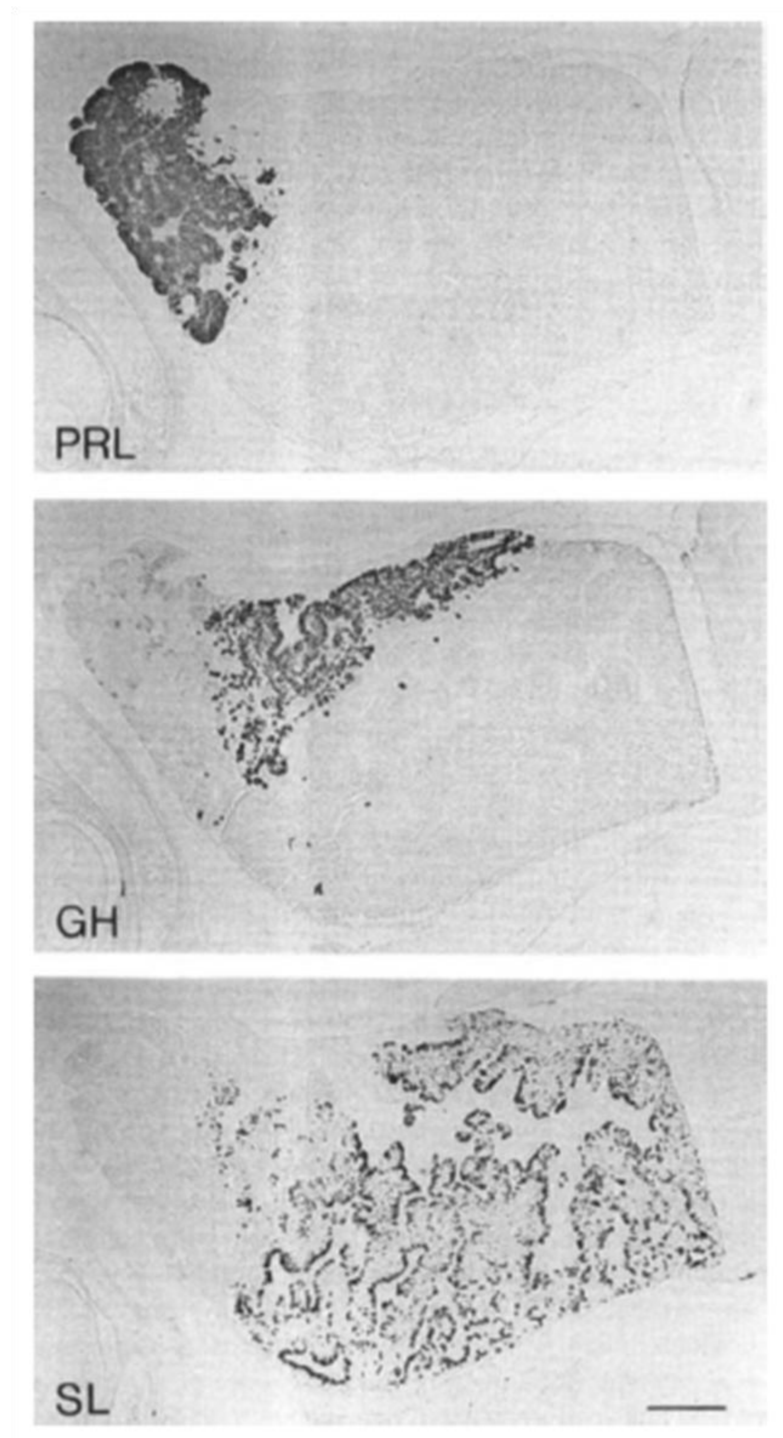


Figure 1: Immunocytochemical localization of PRL, GH and SL in fish pituitary. PRL, GH, and SL producing cells are located in different regions of

the pituitary: in the *rostral pars distalis*, *proximal pars distalis* and *pars intermedia*, respectively. Bar = 200 μm. (Kaneko, 1996)

SL, a fish and tetrapods specific pituitary hormone belonging to helical cytokines superfamily, was isolated for the first time from atlantic cod (Ono *et al.,* 1990). Although SL sequences were recently found on some tetrapods as two species of salamander, the axolotl and the red-spotted newt (Ocampo Daza and Larhammar, 2018), there is still no evidence of SL expression in other than fish species (Fig. 2). SL genomic structure was described in *Oncorhynchus keta* (Takayama *et al.,* 1991) and *Sparus aurata* (Astola *et al.,* 2004) being similar to GH and PRL, with 5 exons and like the other members of this superfamily regulated by Pit-1, a pituitary master regulator (Rand-Weaver *et al.,* 1992; Lopez *et al.,* 2006). SL is expressed in neurointermediate lobe of pituitary gland (Rand-Weaver *et al.,* 1992; Lopez *et al.,* 2006). SL function is still not fully understood but a mutation on SLα in a medaka mutant, color interference (*ci*), shows skin pigmentation problems and lipid imbalance suggesting a possible role in skin color regulation and lipid metabolism (Vega-Rubín de Celis *et al.,* 2003). Also SL was linked to many different physiological functions such as smoltification (Rand-Weaver and Swanson, 1993), acid base balance (Kakizawa *et al.,* 1996), lipid biogenesis (Mingarro *et al.,* 2002), energy mobilization (Rand-Weaver *et al.,* 1993, 1995), gonadal steroid biosynthesis (Planas *et al.,* 1992), metabolism of sodium (Zhu and Thomas, 1995), and calcium (Kakizawa *et al.,* 1993) and gonadal maturation (Planas *et al.,* 1992; Rand-Weaver *et al.,* 1992) (Fig.3). In several bony fish species were described two variants, SLα and SLβ, expressed in different cell populations of NI (Zhu *et al.,* 2004). SLα is highly similar to most SL found in other teleost species, but SLβ only shares a 41.8–49.8% amino acid identity with SL found in most fish species (Valenzuela *et al.,* 2015). These two genes where probably generated by a genome duplication event (Fukamachi and Meyer, 2007).

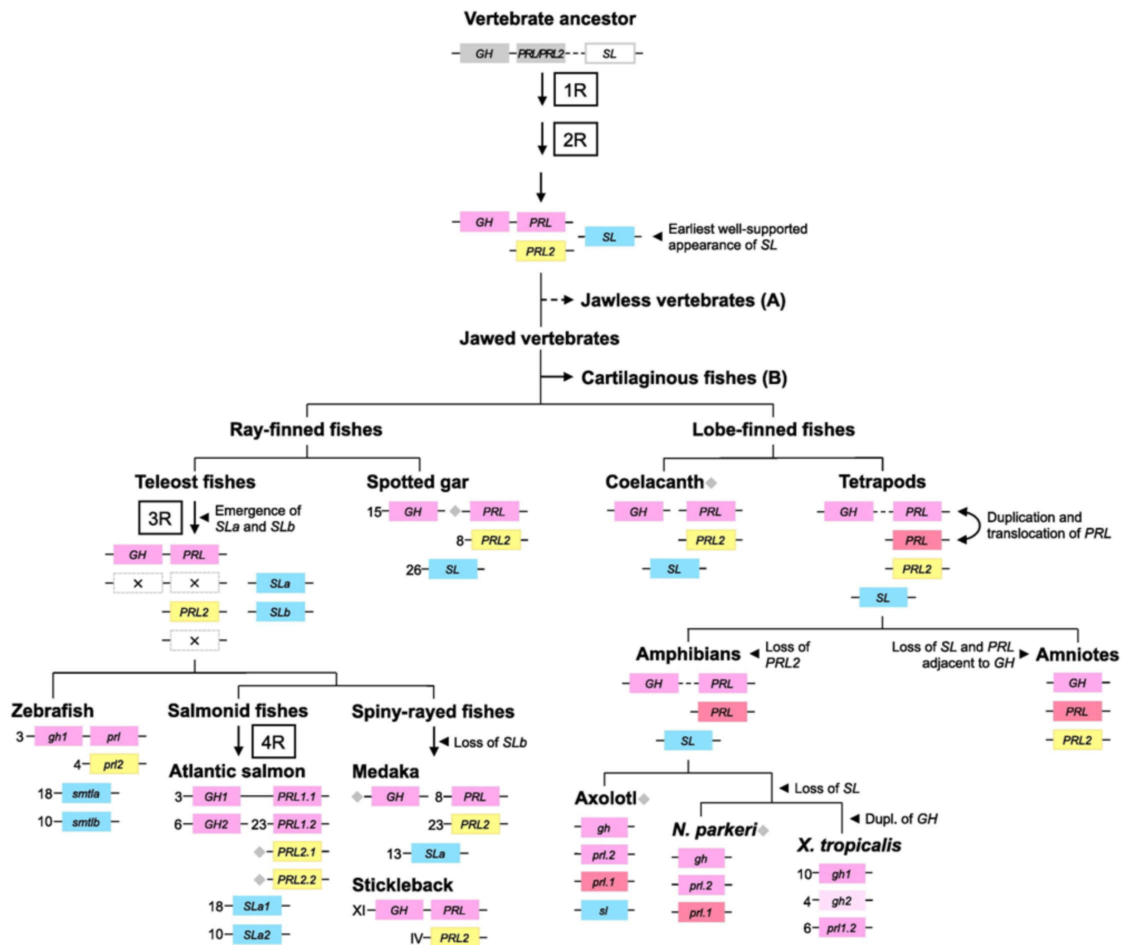


Figure 2: Evolution of GH-PRL superfamily. Evolutionary scenario of the growth hormone family. Based on chromosomal location data and phylogenetic analysis of 105 GH, PRL, PRL2 and SL genes from 28 vertebrate species, 15 of which are shown here (adapted from Ocampo Daza and Larhammar, 2018).

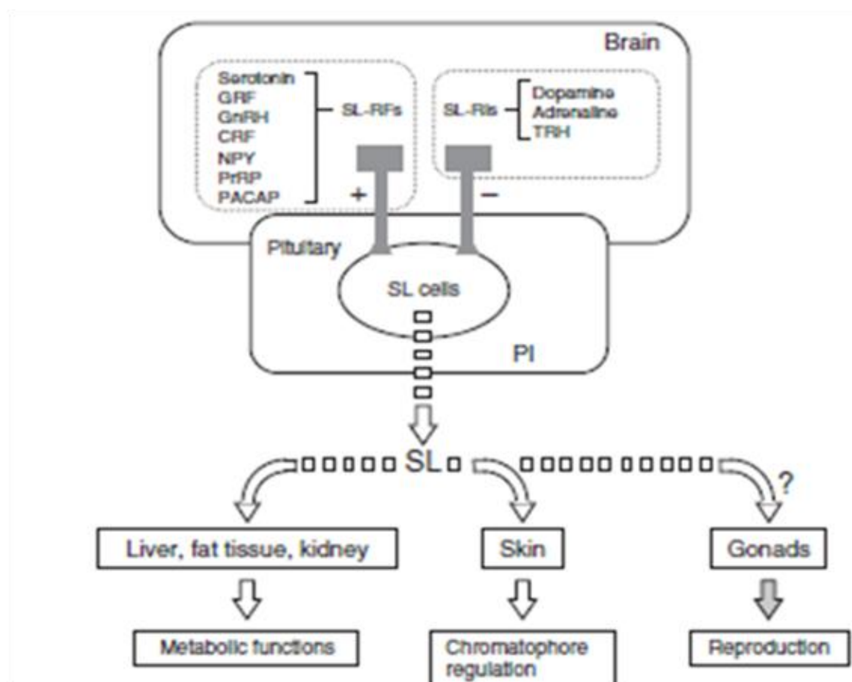


Figure 3: Scheme of SL pathway. SL release from SL-producing cells is under multifunctional control of various neuropeptides and neurotransmitters. These stimulatory/inhibitory regulators from the brain are delivered to SL-producing cells in the PI by direct innervation from the hypothalamus. SL-RFs, somatolactin-releasing factors; SL-RIs, somatolactin-release inhibitors; GRF, growth hormone-releasing factor; GnRH, gonadotropin-releasing hormone; CRF, corticotropin-releasing factor; NPY, neuropeptide Y; PrRP, prolactin-releasing peptide; PACAP, pituitary adenylate cyclase activating polypeptide; TRH, thyrotropin-releasing hormone; PI, *pars intermedia*. (Kawauchi *et al.*, 2009)

Duplicated genomes of bony fish are an excellent model to understand differential expression of duplicated genes in vertebrae. Duplicated genes in eukaryote have a relative short half life, in evolutive time scale, of several million years (Lynch and Conery, 2000) and have a chance of 50% to 92% of being lost (Wagner, 2001). Different models have been proposed to explain for the functional divergence and retention of duplicate genes but the underlying mechanisms are poorly understood.

The theory that gene duplication events are the main source of genetic novelty leading to speciation (Ohno, 1970), has gained wide acceptance (Lynch and Conery, 2000; Gu *et al.,* 2003; Blanc and Wolfe, 2004; Li *et al.,* 2005). According to this theory after a duplication event, one daughter gene keeps the original function, while the other one, in the most of cases, is eliminated, or, in the minority of cases, survives by gaining a new function. This hypothesis, called Duplication-Retention-Non/Neofunctionalization has been the subject of intensive debate (Taylor and Raes, 2004). Under this model, four possible fates await a duplicated gene: Both copies persist, one copy is deleted, one copy functionally diverges or one copy becomes a pseudogene.

The first option is the less probable because complete redundancy is not favored by evolution (Kitano, 2004), retention of two copies will depend on the degree of redundant functions. This is a sort of protection mechanism where any damage on one copy can be buffered by activity of the other gene copy (Fisher, 1935). It has also been suggested that copies may be maintained to prevent a dosage imbalance, in gene expression that would eventually trigger the engagement of very expensive cellular pathways, like alleviating mechanisms of dosage-compensation. Another possible explanation for the preservation of two copies of the same gene is the case of gene products

needed in significant amounts, for example ribosomal RNA or histones, that seems to be a common strategy of prokaryote organisms (Otto and Whitton, 2000).

The most probable outcome of duplication is that one copy of the gene enters a process called non-functionalization, in which one copy is deprived of its function and is eventually lost. Such a mechanism appears not to contribute in terms of adding novelty and variation for evolution as it leads to the restoration of both pre-duplication genotype and phenotype (Ohno, 1970).

The majority of the duplicated genes are not completely deleted from genome but degenerate into pseudogenes (Jacq *et al.,* 1977). Although these DNA sequences display characteristical structural elements of expressed genes, like exon-introns and promoter regions they do not seem to execute any function. Pseudogenization may be triggered by changes in genetic background or environment variation.

If after duplication one copy of the gene becomes functionally redundant this sequence will be free to accumulate mutations without any lethal consequence for the organism. Occasionally, such mutations can by chance derive into a new function, allowing it to be considered as a new gene, this process is called Neofunctionalization (Ohno, 1973). Neofunctionalization means the retention of both gene copies and confers an advantage as both genotype and phenotype can be altered.

However, three observations on genome-wide duplication events are contradictory to this theory, first a higher proportion of the duplicated genes retained than expected by chance alone, second, nucleotide substitution patterns reflective of purifying selection on both copies of the duplicated

genes, and third, a relative paucity of null allele for loci that have avoided nonfunctionalization. These findings lead to a new current of thought. The first theory postulated that the gene's ancestral functions are partitioned between the two daughter genes, resulting in a recovery of the original phenotype accompanied by a variation in the genotype (Orgel, 1977). This model proposed that sharing of an original function occurred as the way diversification takes place in multigene families. This proposal referred as "subfunctionalization" took shape in the duplication-degeneration-complementation (DDC) mathematical model (Force *et al.,* 1999; Lynch and Force, 2000). DDC attributed complementary degenerative mutations in regulatory elements for controlling the expression of duplicated genes, leading to portioning of ancestral gene functions. If any selective pressure acts on this mechanism, the evolutionary advantage of subfunctionalization is presumably to allow individual optimization of the newly separated functions, although it could simply represent the most likely outcome for complexly regulated genes.

As both neofunctionalization and subfunctionalization alone adequately explain the genome-wide patterns of gene expression for duplicate genes, a more complex scenario known as "sub-neo-functionalization" has also been proposed combining elements of neofunctionalization and DDC models (He and Zhang, 2005). The two paralog expression patterns will diverge as a result of modifications in cis-regulatory elements, which allow them to bind different sets of transcription factors. According to this theory, a large proportion of duplicate genes number of paralogs have endured rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization (He and Zhang, 2005). More recent work on duplication and divergence of GAL genes and their contribution to genetic novelty provided experimental evidence in support of this model (Hittinger and

Carroll, 2007). These findings do not exclude the possibility that a minority of duplicate genes evolve by pure subfunctionalization or pure neofunctionalization. However, it was found that subfunctionalization occurs rapidly after gene duplication, whereas neofunctionalization is a lengthy process that continues even long after duplication. Thus, the short-term retention of duplicate genes in the genome is primarily due to subfunctionalization, consistent with a much higher rate of degenerate mutations than beneficial mutations (Walsh, 1995; Lynch and Force, 2000).

Promoter regions appear to be of particular relevance in the study of duplicate genes and have been the focus of several investigations. Regulatory regions of duplicated genes have been reported able to diverge, allowing new functions as well as changes in expression levels (Louis, 2007). As a result of these changes, new or complementary organism functions could arise from differential regulatory mutations (Hellsten *et al.,* 2007). The well-studied example of Hox1b duplicates in zebrafish suggests that observed degeneration of discrete and complementary cis-regulatory elements might underlie the subfunctionalization of expression patterns (Prince and Pickett, 2002). In yeast (*Saccharomyces cerevisiae*), recently duplicated genes were shown to rapidly gain transcription factor binding sites after duplication (Tsai *et al.,* 2012). Also was found that duplicated gene pairs vary greatly in their cis-regulatory element architecture, resulting in changes in regulatory network connectivity (Long *et al.,* 2016).

Whole-genome duplicates have approximately twice as many footprints in their promoters left by potential regulatory proteins than do tandem duplicates. The footprints, in turn, result in more regulatory network connections. In carp evolution one of the latest whole genome duplication

(WGD) occurred about 8-14 MYA (Omori y Kon, 2018). Whole-genome duplicates and other genes, forming denser, more complex regulatory networks than shown by tandem duplicates (Arsovski *et al.,* 2015). In a study of the particular effects of Whole-genome duplication on network architecture in yeast, evidence was found for a partitioning of gene expression among duplicate genes produced by Whole-genome duplication. In particular, it seems that one member of a paralogue pair created by Whole-genome duplication is often assigned to stress response pathways, whereas the other paralogue does not act in these pathways. Also, following a genome duplication the number of interactions between proteins was transiently quadrupled, after which a process of interaction loss simplified the network. Interactions can also be gained during this time, but this process is generally considered to be rarer (Fig. 4) (Conant and Wolfe, 2006).

All these findings provide a conceptual framework for the present study because a large portion of the carp genome, including SL genes, is believed to derive from evolutionarily recent polyploidy events (Xu *et al.,* 2014).

Although genetics determines the endocrine phenotypes, and the tight regulation by transcription factors allows an adaptation to different stimuli, it cannot explain by itself the great variability and reversibility of the system in response to environmental changes (Rodin and Riggs, 2003; Rapp and Wendel, 2005; Rodin *et al.,* 2005). Current evidence suggests that epigenetic mechanisms, through inheritable but reversible changes that are not based on alteration of the nucleotide sequence, constitute the nexus between genetics and environmental modulation of endocrine functions. Epigenetic mechanisms, including DNA methylation, modification and histone and microRNA variants, divide the genome into active and inactive domains based

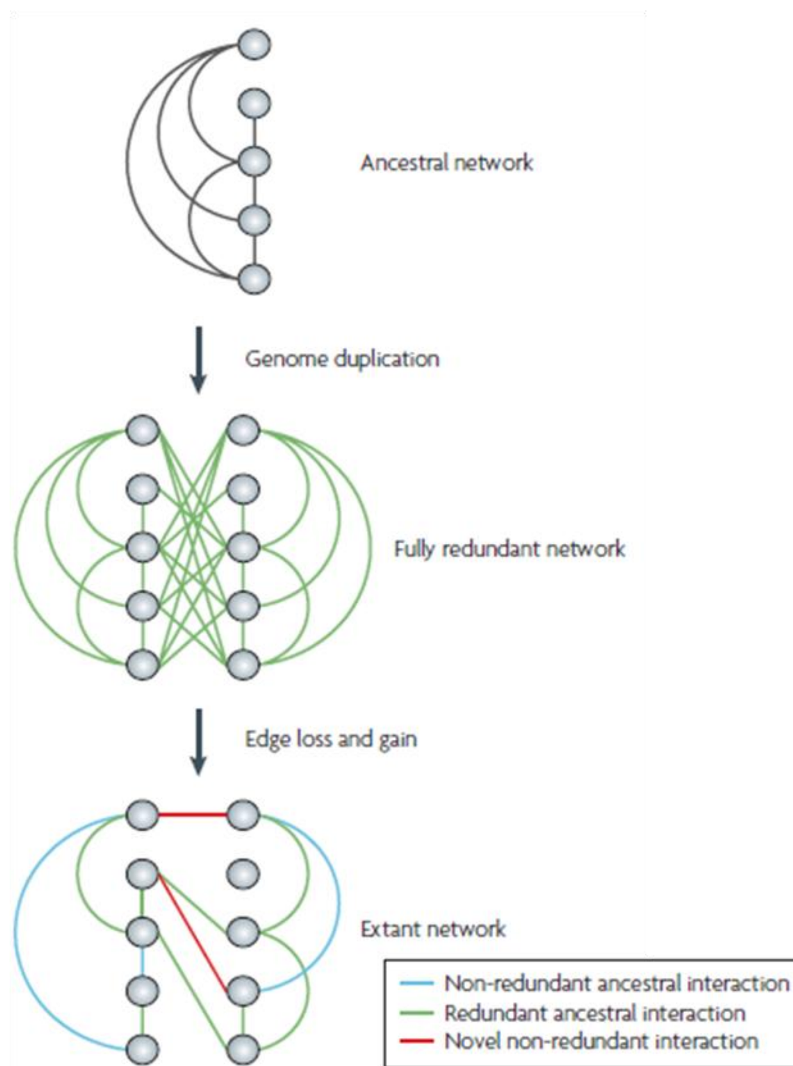


Figure 4: Hypothetical example of network evolution following a genome duplication. Proteins are represented as circles, with interactions between a pair of proteins represented as joining lines. Note that it was assumed that the ancestral network is known (which is generally not true for real networks), a fact that allowed to distinguish interactions surviving from the duplication from novel interactions. Figure is modified from (Conant *et al.,* 2008)

on endogenous, exogenous environmental changes and developmental stages, generating phenotypic plasticity (Jaenisch and Bird, 2003; Németh and Längst, 2004). The most studied histone modifications are posttranscriptional modification of specific histone amino acid residues by acetylation/deacetylation, methylation/demethylation and phosphorylation/dephosphorylation (Bannister and Kouzarides, 2011). These chemical modifications affect the interactions of basic, positively charged histone proteins with negatively charged DNA, resulting in different chromatin states, a relaxed accessible euchromatin and a condensed heterochromatin.

These states are linked to active and inactive gene expression by regulating the accessibility of DNA to transcription factors and ncRNAs involved in transcriptional regulation of genes (Allis and Jenuwein, 2016). Within the major groups of histone modifications, histone acetylation dynamics that largely affect the ε-amino group of lysine (K) residues at the N-terminal of H3 and H4, including a classical heterochromatin H3K4me3 and eucromatin mark H3K9me2 (Zhao and Garcia, 2015). Recent studies suggest that specific histone marks are highly conserved between zebrafish and mammals (Cunliffe, 2016) and are linked to permissive or inhibitory chromatin state for gene expression at least in part by regulating accessibility to transcription factors (Joseph *et al.,* 2017).

DNA methylation occurs to methylate cytosines which are mostly, but not exclusively located in the context of genomic CpG dinucleotides (Ramsahoye *et al.,* 2000; Koganti *et al.,* 2017). Such DNA methylation can subsequently modify gene transcription as, for instance, methylation of CpG sequences at transcription start sites has been associated with long-term silencing (Jones,

2012). However, DNA methylation also occurs outside of promoter regions, and relationship between DNA methylation and gene silencing has proved to be challenging to reveal (Ambrosi *et al.,* 2017). Studies in teleost fish have revealed functional similarities between zebrafish and mammalian model systems in DNA methylation dynamics (Goll and Halpern, 2011). However, while methylation of promoter regions is often claimed to be inversely related to gene expression, methylation status of DNA sequences within the genes (especially exons) showed even better correlation with gene expression in zebrafish (McGaughey *et al.,* 2014).

Recent studies gave strong support to the idea that epigenetic divergence of duplicate genes affects gene expression and, ultimately, functional divergence of duplicate genes. It is interesting that gene-body DNA methylation does not show a discernible relationship with evolutionary age compared with promoter methylation (Keller and Yi, 2014). Together these evidences support that epigenetic modifications are important facilitators of duplicated gene evolution owing to their effect on functional divergence strengthening the relationship between genomes and epigenomes.

These evolutionary mechanisms have undoubtedly favored the ability of organisms to adapt to changes in the environment, however, the endocrine system of vertebrates is susceptible to disruption by environmental chemicals such as anthropogenic industrial products and as well as by natural compounds (Diamanti-Kandarakis *et al.,* 2009; Gore *et al.,* 2015; Carnevali *et al.,* 2018). These compounds are named endocrine disrupters (EDs) and defined by the International Programme on Chemical Safety as an "exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its

progeny or (sub)populations". Concern has been raised by the consequences that they might have on both wildlife and the human population (Colborn *et al.,* 1993; Sonnenschein and Soto, 1998). Freshwater ecosystems are particularly vulnerable to the presence of EDs since the proximity to the sources of pollution and the low dilution factor of these waters makes the inhabiting fauna highly exposed. Estuaries and their associated marine waters are also affected by high levels of estrogenic compounds and the effect on the open sea, although still unknown, cannot be ignored.

Some of these EDs can mimic the effects of estrogens. Estrogen is the primary female hormone, it is responsible for the development and regulation of the female reproductive system and secondary sex characteristics. Like all steroid hormones, estrogens readily diffuse across the cell membrane. Once inside the cell, they bind to and activate estrogen receptors (ERs) which in turn modulate the expression of many genes (Nussey and Whitehead, 2001). However, from natural estrogens or synthetic estrogens with pharmaceutical use disposed through domestic waste water to environmentally persistent man-made chemicals and sometimes their major degradation products, as well as xenoestrogens from pulp and paper industries, might be leading to subtle, but potentially very serious detrimental effects on aquatic environment (Hewitt *et al.,* 2008; Gore *et al.,* 2015). Most of the estrogenic chemicals discussed above are lipophilic and hydrophobic and, hence, have a strong tendency to bioconcentrate and bioaccumulate in aquatic organisms, both plants and animals. Additionally, estrogen caused a considerable reduction in fish biomass and in consequence interrupt the aquatic food chain (Hallgren *et al.,* 2014; Carnevali *et al.,* 2018). Also, in higher concentrations estrogen does have severe deleterious effects on other forms of aquatic life. For example, 17β-estradiol at 10 ng/L dramatically affected the heart function of bullfrog

tadpoles (Salla *et al.,* 2016). Despite that plants synthesize phytoestrogens, they also can take up animal-derived estrogens both actively and passively. Estrogens lipophilic properties facilitate easy passage through plant membranes and therefore can accumulate in both roots and shoots (Bircher, 2011). In a concentration dependent manner, estrogens can stimulate or inhibit plant growth and development (Pocock and Falk, 2014). Estrogen hormones play a key role in human biology and physiology. They help regulate reproduction, cognitive behavior, cardiovascular function, bone density, behavior and digestive system. Steroidal estrogens in food and water can also affect reproductive development, induce premature menopause and cause virilization in young women. On the other hand estrogens were clearly shown to be involved in the decline of sperm counts and disorders of the male reproductive system and feminization of men (Bolong *et al.,* 2009; Sumpter and Jobling, 2013). Besides the estrogenic effect on reproduction, the immune system and metabolism is also affected (Sirotkin and Harrath, 2014).

These evidences suggest that it is crucial to ensure that both estrogens from human and animal waste and phytoestrogens are not consumed in food and water at levels above the accepted because this could trigger serious health effects. Although worth efforts worldwide, comprehensive EDs monitoring and risk-assessment still require improvements, for this is vital to understand how estrogens affect animal physiology. Exposure of male fish to different concentrations of both natural and manmade estrogens has shown very pronounced dose-response effects (Bromage and Cumaranatunga, 1988) and has also shown that male fish are very sensitive to estrogens present in the water. In the carp, a vitellogenin response has been observed at 0,1 ng/L of 17β-ethinylestradiol after only a relatively brief exposure (Smeets *et al.,* 1999). Effluent from industrial and domestic wastewater contains a chemical, or

more likely a combination of chemicals, which are absorbed by fish and "feminize" the fish, modifying sexual behavior of fish (Coe *et al.,* 2010; Reyhanian *et al.,* 2011; Filby *et al.,* 2012), disrupting reproductive capacities of fish *(Länge et al., 2001; Nash et al., 2004; Pawlowski et al., 2004; Fenske et al., 2005; Carnevali et al.,* 2018) and eventually generating an adversely impact recruitment of fish populations (Kidd *et al.,* 2007; Lange *et al.,* 2011). In addition to the reported effects, estrogenic exposure can also cause alterations on the xenobiotic metabolizing capacity of fish (Arukwe and Goksøyr, 1997; Solé *et al.,* 2000). All of these changes might contribute to the hormonal imbalance observed in pollution-exposed organisms.

Recently we described that estrogen-treated male carp showed a significant increase of *sl*β expression, but no significant change in *sl*α expression (Fig. 5).

This suggests that both genes are differentially regulated and might reflect a subfunctionalization or a subneofunctionalization of *sl*α and *sl*β genes in response to estrogen. However, the mechanisms involved in this phenomenon are still unclear (Valenzuela *et al.,* 2015).

Therefore, comparative analyses of both *sl* genes that have evolved for millions of years in the same cell functioning in the adult organism, pinpointing changes on genetic and epigenetic level will contribute novel insights in underlying mechanisms for differential gene regulation in the complex interplay of genome and environmental cues.

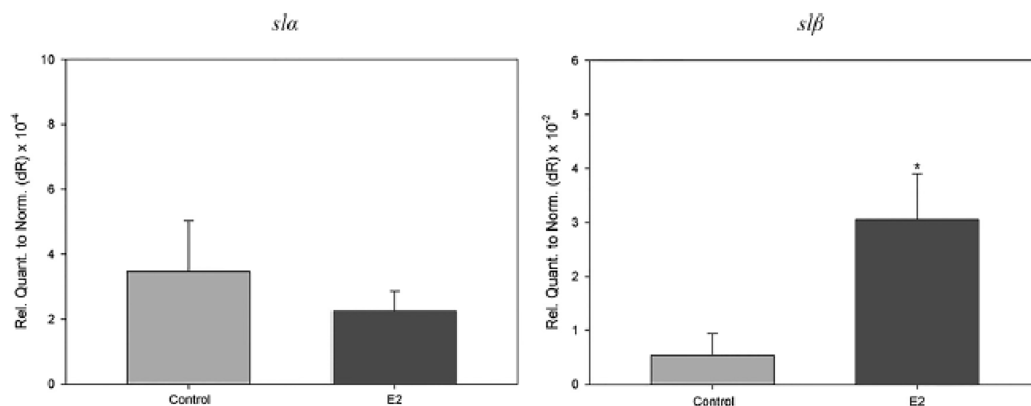


Figure 5 Estrogen effect on gene expression of hypophyseal and hypothalamic factors in male carp. mRNA levels of *sl*α and *sl*β in pituitary measured using Quantitative reverse transcription PCR (RT-qPCR) and represented as relative quantification to normalizer gene. Analyses were performed in duplicate and all data were normalized for β-actin gene expression. Graphs depict relative quantification to normalizer gene (dR) from four individual adult male carp with specific treatment (n = 4) and corresponding control group (n = 4). Bar indicates standard deviation. (*) Applies to significant difference. Student's t-test, P < 0.1 was considered significant difference between controls and estrogen treated animals (adapted from Valenzuela *et al.,* 2015).

Together these evidences lead to postulate the following working hypothesis:

"The differential expression of *sl* genes in response to estrogen is related to genetic elements Pit-1, ER and epigenetic effects in *Cyprinus carpio* pituitary."

For which the following objectives were proposed in order to solve the hypothesis:

General objective:

Study the regulation of SL expression, characterize gene elements of both *sl* genes and epigenetic effects in response to estrogen in carp pituitary.

Specific objectives:

1 Reveal regulatory elements in the sequences of *sl*α and *sl*β genes in *Cyprinus carpio* and identify conserved and divergent elements

2 Analyze the binding sites of Pit-1 and ER transcription factors and the influence of ER binding sites on estrogen responsiveness

3 Characterize the methylation profile of *sl*α and *sl*β in response to estrogen

4 Determine chromatin structure *in vivo* in *sl*α and *sl*β in response to estrogen

4. MATERIALS AND METHODS

4.1. MATERIALS

4.1.1. Equipment and Instruments

- Class II Biological Safety cabinet: Nuaire NU425-400E.

- Centrifuges: Sigma 2-16 PL, Sigma 1-14 Microfuge, Boeco C-28A. 35, High Speed Refrigerated centrifuges Hitachi Himac CR22-GII, Preparative Ultracentrifuge Hitachi Himac CP20-WX.

- Electrophoresis systems: BioRad Mini-Protean III,

- Electroblotting system: Labnet Enduro modular Vertical gel system.

- Freezer -20ºC: Cónsul.

- Freezer -80ºC: Ult-Freezer 560 liters.

- Horizontal gel box electrophoresis: Labnet Enduro.

- Incubator: Zhicheng ZSD 1270.

- Incubator Shaker: Zhcheng, ZHWY-200B.

- Microwave oven: Somela Faney WT1700.

-Microscope: LW-Scientific I4 Series, Olympus CKX41, Inverted Microscope.

- Micropipettes: Gilson PIPETMAN.

- Power supplies: Biorad Power PacTM Universal Power Supply, Enduro E0303 model 300V power supply.

- Potter-Elvehjem Tissue Grinder 5mL: Weathon Science.

- Platform Shaker: Heidolph polymax 1040.

- Precision Balance: Sartorius TE4101.

- pH meter: Benchtop precision pH meter WTW InoLab pH720.

- Real-Time PCR System: Stratagene MX 3000P.

- Refrigerator: Fenza.

- Spectrophotometer: Thermo Scientific Evolution 60, GE Healthcare NanoVue.

- Thermoblock: Labnet, AccuBlockTM Digital Dry Baths.

- Thermal Cycler: Labnet, MultiGene™ OptiMax Thermal Cycler, Eppendorf, Mastercycler personal.

- Ultrasonic processor Cole Parmer, CPX130PB.

- UV transilluminator: Syngene, INGENIUS.

- Vortex mixer: Brarnstead International, MAXIMIX II VORTEX MIXER.

- Water Bath: N-Biotec, NB-301.

4.1.2. Solutions

• **Buffer SB 1X:** 10 mM sodium hydroxide, pH adjusted to 8.5 with boric acid.

• **Buffer MOPS:** 20 mM MOPS pH 7.0, 2 mM sodium acetate, 1 mM EDTA pH 8.0.

•**Tris/Borate/EDTA (TBE) buffer:** 90 mM Tris-HCl, pH 8.0, 90 mM boric acid, 2 mM EDTA

• **Buffer TAE:** 40 mM Tris-acetate, 1 mM EDTA**.**

• **PBS**: 136.89 mM NaCl, 2.68 mM KCl, 10.14 mM Na_2HPO_4, 1.76 mM KH_2PO_4 pH 7.4.

• **Luria Broth (LB) medium:** 1% (w/v) Bacto Tryptone, 1% (w/v) NaCl 0.5% (w/v) Bacto Yeast Extract. pH to 7.0, sterilized by autoclaving

• **Electromobility shift assays**

o **Annealing buffer** 20 mM Tris-HCl, pH 7.6, 50 mM NaCl, 2 mM $MgCl_2$

• **Polyacrylamide gel electrophoresis:**

o **Loading buffer SDS-PAGE:** 62.5 mM Tris-HCl, pH 6.8; 20% glicerol, 2% SDS, 5% β-mercaptoetanol 0, 002% bromophenol Blue.

o **Running buffer SDS-PAGE:** 25 mM Tris, 192 mM glycine, 0.1% SDS; pH 8.3.

o **Stacking gel buffer SDS-PAGE 4x:** 0.5 M Tris-HCl pH 6.8, 0.4% SDS.

o **Resolving gel buffer SDS-PAGE 4x:** 1.5 M Tris-HCl pH 8.8, 0.4% SDS.

• **Coomassie Staining:**

o **Fixing solution:** 50% methanol and 12% glacial acetic acid.

o **Coomassie blue staining solution for polyacrylamide gels:** 0.3% Coomassie Brilliant Blue R-250, 50% methanol and 10% glacial acetic acid.

o **Destaining solution:** 30% methanol and 7% glacial acetic acid.

• ChIP:

o **Cross-Linking solution:** 1% Formaldehyde, 20 mM Hepes-NaOH, 40 mM NaCl, 0.2 mM EGTA, 0.4 mM EDTA in 1x PBS

o **Lysis buffer:** 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH8.0.

o **Wash buffer:** 50 mM Hepes (pH 7.6), 1 mM EDTA, 1% NP-40, 0,7% Na-Deoxycholate 0.5 M LiCl.

o **TE buffer:** 10 mM Tris-HCl (pH 8.0), 1 mM EDTA

o **Elution buffer:** 50 mM Tris (pH 8.0) 10 mM EDTA, 1% SDS

• Protein Purification

o **Column Lysis Buffer** 10 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0

o **Column Wash Buffer** 20 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0

o **Column Elution Buffer** 250 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0

4.1.3. Software and on line tools

- http://www.ncbi.nlm.nih.gov/ National Center for Biotechnology Information (NCBI).
- Geneious 7

4.1.4. Biological material

4.1.4.1 Animals

Adult male carp (*C. carpio*) weighing between 1 and 2 kg were captured from surroundings of Valdivia and maintained for 2 weeks in a pond near Valdivia with flow through spring water at environmental temperature and photoperiod.

4.1.4.2 Bacterial strains

According to Table 1.

strain	genotype	use
XL1 Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZDM15 Tn10 (Tetr)]	Cloning, plasmid propagation, blue/white screening
Rosetta (DE3)	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysSRARE (CamR)	Recombinant protein expression

Table 1: Bacterial strains used for cloning and protein expression

4.2. METHODS

4.2.1 Animals treatment

Selected male carp were injected intraperitoneally with 0.5 mg/kg 17ß-estradiol dissolved in vegetable oil/ethanol (9:1), other four individuals with vehicle only for three consecutive days and sacrificed on the fourth day as described earlier (Figueroa *et al.,* 1994). The performed experiments complied with the Guidelines of the Comisión Nacional de Ciencia y Tecnología de Chile (CONICYT) and the "Universidad Austral de Chile" for the use of laboratory animals.

4.2.2 Molecular cloning

All work was performed using standard protocols and handled with standard precautions, according to the guidelines of biosafety level 1 (S1) laboratory work. Bacteria were cultured in LB medium supplemented with antibiotics. Medium was stored at room temperature, antibiotics at -20 °C. Bacteria were grown at 37 °C with agitation in liquid medium or without agitation on solid medium plates.

4.2.2.1 Isolation of plasmid DNA

Isolation of plasmid DNA from 5 ml liquid bacterial culture was performed using the EZNA® Plasmid Mini Kit I, (Q-spin). If a higher quantity of endotoxin-free, transfection-grade plasmid DNA was required, 50 ml of liquid LB medium was inoculated and the PureLink® HiPure Plasmid Filter Midiprep Kit was used for isolation. Both Kits were used according to the manufacturer's instructions.

4.2.2.2 Determination of nucleic acids quantity and purity

Quantification of DNA and RNA concentration was carried out using UV/Vis spectrophotometric (Thermo Scientific Evolution 60, GE Healthcare NanoVue). At 260nm for DNA one OD corresponded to 50 μg/ml and RNA 40 μg/ml. Samples with an A260/A280 ratio of 1.8 or higher are considered to be free of protein contamination and aromatic substances, while A260nm/A280nm of > 2.0 indicates contamination other nucleic acid. In addition, the A230/A260 ratio hints to possible contamination with organic solvents often used in DNA purification, such as ethanol, phenol or chloroform. NanoVue measurement was used for all standard molecular biology applications, such as cloning, plasmid purification or DNA extraction.

4.2.2.3 Restriction enzyme digest

All restriction enzyme digests were carried out according to the manufacturer's instructions on enzyme combinations, buffer usage, BSA supplementation and incubation temperature and times (see http://www.neb.com/nebecomm/default.asp). High Fidelity (HF) endonucleases were used whenever possible. Preparative digests e.g. for subsequent ligation reactions were performed with 10 μg of DNA and 50 U of the respective enzyme(s) in a total reaction volume of 50 μl at 37 °C for 2 – 4 h. For all restriction digests, a final concentration of > 5% of glycerol was avoided because this might cause star activity. As restriction enzymes are commonly stored in 50% glycerol, the added enzyme volume should not comprise more than 10% of the total reaction volume. Digests were resolved on 1% agarose gels containing 0.01% SYBR®Safe in 1x TAE buffer as described in 4.2.2.15.

4.2.2.4 Polymerase chain reaction (PCR)

PCR was used for amplification of a template, introduction of restriction enzyme cutting sites for subsequent cloning and/or construction of deletion mutants. The regular PCR reaction was performed using the following temperature cycles: Initial denaturation 95°C during 5 min, follow of 35 cycles of denaturation 95°C for 30 s, annealing *50-65°C for 30 s, extension 72°C for 20 s, and Final extension of 10 min at 72°C. The PCR reaction was performed in a total volume of 20 μL containing 0.5 U de GoTaq flexi DNA polymerase, 1x Green GoTaq Flexi Buffer, 3 mM $MgCl_2$, 0.2 mM mix dNTPs, 0.5 μM of each primer, 100 ng template DNA, in DEPC water. For high fidelity or long fragment PCR, e.g., reporter assay constructs, or amplification of Pit-1 coding sequence, Phusion® High-Fidelity DNA Polymerase from NEB was used, according to the manufacturer's instructions.

* Annealing temperature depends on the primers used, typically between 55 and 65 °C. Primer were annealed at + 3 °C above the lowest TM (TM as suggested by the oligonucleotide manufacturer). Difficult templates or long primers might require prior testing of annealing temperatures by running a gradient PCR. The PCR products were analyzed on 1% agarose gels containing 0.01% SYBR®Safe (see 4.2.2.15).

4.2.2.5 Purification of PCR products

Purification of PCR products from primers and dNTPs and de-salting was performed with the QIAquick® PCR Purification Kit according to the manufacturer's instructions.

4.2.2.6 Purification of DNA by phenol/chloroform

To separate nucleic acids from proteins, phenol/chloroform extraction was used. For DNA extraction phenol at neutral pH was used. One volume of phenol/chloroform/isoamyl alcohol (50:49:1) solution was added to the sample and mixed vigorously by vortexing. Phases were separated by centrifugation for 5 min at 13.000 xg at room temperature. The upper, aqueous phase containing the DNA was transferred carefully to a new tube without touching the protein layer. A second round of phenol/chloroform-isoamyl alcohol extraction was performed as described above. In order to remove traces of phenol, the aqueous phase was extracted with one volume of chloroform:isoamylalcohol 49:1, mixed and centrifuged for 5 min at 13.000 xg at room temperature. The supernatant was transferred to a new tube. DNA was precipitated using acetate and two volumes of ice-cold ethanol. Sodium acetate (0.1 volumes, 3 M, pH 5.2) was used for Cloning and Bisulfite assays. After addition of acetate and ethanol, samples were incubated for 10 – 30 min on ice. Next, the tubes were centrifuged for 15 min at 13.000 xg and 4 °C. The supernatant was aspirated carefully and the pellet was washed with one volume of 70 % ethanol and centrifuged another time for 10 min at 13.000 xg and 4 °C. The supernatant was aspirated carefully and the pellet was air-dried for approximately 10 min. Finally, DNA was dissolved in the appropriate amount of water or Tris-HCl buffer (10 mM, pH 8.5) and stored at – 20 °C.

4.2.2.7 Inverse PCR

This technique allowed to obtain sequences located towards the 5 '-end of the coding region of *sl* previously identified and sequenced. This approach involved the following: A restriction enzyme site was located in the sequence near the 5'-end of the known sequence, the genomic DNA was digested with

this enzyme, the assay was diluted and the individual fragments were religated. In the circular DNA mixture the 5'-region was amplified with a pair of starters oriented in the opposite direction in the region most towards the 5'-end of the known region. The amplification product represented the 5'-end of the known sequence and the sequence until the next cut of the restriction enzyme with which the genomic DNA was digested and in this way progress was made towards revealing the promoter sequence.

For the *sl* promoter, this approach of the inverse PCR was carried out with two different restriction enzymes. Two pairs of primers were designed within the coding region of *slα*, before the cleavage position of the *Sac*I enzymes and another for the *Xba*I enzyme, pointing in the opposite direction to the common oligonucleotides used in conventional PCR, i.e. the sense oligonucleotide in the 3'-end direction and the 5'-end antisense (Ochman *et al.,* 1988). First step was digest two aliquots of genomic DNA, between 0.5 to 1 μg, with *Sac*I and *Xba*I, respectively, incubating 10 μL of genomic DNA, 2 μL of 10x restriction enzyme buffer (NEB), 1 μL of the restriction enzyme (NEB) and 7 μL of H_2O, at 37°C overnight, to ensure complete digestion. To check if the digestion was effective, it was controlled by fractionating an aliquot on a 2% agarose gel to see the appearance of a smear of the fully digested genomic DNA. The rest of the digestion was diluted to favor the ligation of the fragments with itself, performing the ligation reaction under the following conditions: at 20 μL of the digestion reaction, 200 μL of H_2O, 25 μL of buffer were added. 5x ligation (NEB), 5 μL DNA Ligase T4 (3 U/μL, NEB) and incubated at 4°C overnight. Then, 1 μL of this reaction was used as template for the first PCR reaction, using the same protocol detailed in point 4.2.2.4 with the difference that 1 min of extension was used, using oligonucleotides of Table 2. A second PCR reaction (nested) was performed

with 1 μL of the first PCR reaction as annealed with the same primers in a second round of PCR reaction, under the same conditions. The products were cloned into the pGEM-T vector, as described in section 4.2.2.12, and then sequenced by Sanger Method (Macrogen, South Korea). The sequences obtained were manually cured and specificity was confirmed by alignments with the coding sequence of *sl*α and *sl*β using the Geneious 7 program.

Name	Sequence	Target gene	Use
cSLAinva	GTCTCGTCTTTGCAGTCCAG	slα	Inverse PCR
cSLAinvs	GCGTCATCCAACATGCAGAG	slα	Inverse PCR
cSLBinva	TTGTGTGGCGTTTGTGCTCT	slβ	Inverse PCR
cSLBinvs	GTAGAGCTGTAGCTTTCTTC	slβ	Inverse PCR
cSLA2inva	GCACTTCAATCTCTCAAGAGAA	slα	Inverse PCR (Nested)
cSLA2invs	GAGGAATGGCGTGAGAAAGC	slα	Inverse PCR (Nested)
cSLB2inva	TCACTGCAGCCCGTGACTGG	slβ	Inverse PCR (Nested)
cSLB2invs	AGAGCTGTAGCTTTCTTCAT	slβ	Inverse PCR (Nested)

Table 2. Oligonucleotides used to amplify promoter region of sl genes

4.2.2.8 5`-Rapid Amplification of cDNA Ends (RACE)

This method is used to extend partial cDNA clones by amplifying the 5'-sequences of the corresponding mRNAs. The technique requires knowledge of only a small region of sequence within the partial cDNA clones in this case the coding sequence of *slα* (Ac. Num. GU434163.1) and slβ (Ac. Num. DQ021542.1). During PCR, the DNA polymerase was directed to the appropriate target RNA by a single primer derived from the region of known sequence; the second primer required for PCR is complementary to a general feature of the target, in the case of 5'-RACE, to a homopolymeric tail added (via terminal transferase) to the 3'-termini of cDNAs transcribed from a preparation of mRNA. This synthetic tail provides a primer-binding site upstream of the unknown 5'-sequence of the target mRNA. The products of the amplification reaction are cloned into a plasmid vector for sequencing.

For this 1 μg of total RNA was retrotranscribed with gene-specific antisense primer, cSLA1a and cSLB1a, (Table 3) according to 4.2.4. The cDNA was then purified, with a commercial kit as Wizard SV Gel and PCR clean-up system (PROMEGA). A PolyA tail was then added to the cDNA, using 30 units of Terminal deoxynucleotide transferase (TdT) and 1 μL of 10 mM dATP, 4 uL of the TdT enzyme buffer in a total volume of 20 μL for 15 min at 37°C followed 3 min at 80°C to denature the enzyme. The result of the reaction was brought to a volume of 450 μL and used as a template for a PCR reaction using an adapter oligonucleotide 1 (GACTCGAGTCGACATCGA$(T)_{17}$), adapter oligonucleotide 2 (GACTCGAGTCGACATCG) in addition to a specific oligonucleotide complementary to the sequence known mRNA to be amplified, cSLA2a and cSLB2a (Table 3), more towards the 5` than the one used for reverse

transcription. The reaction began with a first cycle of 5 min at 94°C, 5 min at 55°C and 40 min at 72°C. Followed by 30 cycles of 40s at 94ºC, 1 min at 55ºC and 3 min at 72ºC. Concluding with a final extension at 72ºC for 15 min. The PCR products were fractionated on a 1.5% agarose gel and the band of the expected size was cut out and purified using the Wizard SV Gel and PCR clean-up system (PROMEGA) kit, then cloned into the vector pGEM- T Easy and subsequently sequenced (Macrogen, South Korea).

Name	Sequence	Target gene	Use
cSLA1a	AAGGCAGCTGAGCTTGTTGC	slα	5`RACE RT
cSLA2a	GAGCAGCTTGAGGAAGGTCT	slα	5`RACE PCR
cSLB1a	GCTCTGGATGTTGAACGGCT	slβ	5`RACE RT
cSLB2a	GGCAGTATCTTCATCTGGACAG	slβ	5`RACE PCR
cSLA3s	TTCTGACTCCGTCCGTCTTC	slα	3`RACE
cSLA4s	GCAACAAGCTCAGCTGCCTT	slα	3`RACE Semi nested
cSLB3s	TGGATTAATCCACTGGTAGA	slβ	3`RACE
cSLB4s	GGTTTGGTGTTGGAAGGAGC	slβ	3`RACE Semi nested

Table 3. Oligonucleotide for amplification of cDNA ends.

4.2.2.9 3`-RACE

Total RNA (1 µg) was retrotranscribed according to 4.2.4 and used as a template for a PCR reaction using forward primer designed in the coding sequence of *slα* (Ac. Num. GU434163.1) and *slβ* (Ac. Num. DQ021542.1) and oligodT as reverse primer, followed by a seminested PCR reaction with a forward primer located near to first primer and again oligo-dT as reverse primer (Table 3). The reaction began with a first cycle of 5 min at 94°C, followed by 30 cycles of 30 s at 94ºC, 30 s at 55ºC and 3 min at 72ºC, concluding with a final extension at 72ºC for 15 min. The PCR products were fractionated on a 1.5% agarose gel and the band of the expected size was cut out and purified using the Wizard SV Gel and PCR clean-up system (PROMEGA) kit, then cloned into the vector pGEM- T Easy and subsequently sequenced (Macrogen, South Korea).

4.2.2.10 Gibson assembly

Vector pGL3 and promoter of *slβ* were amplified, wild type or without ERE using primers of Table 4, in the following order: pGL3s and pGL3a for amplify pGL3 vector backbone, then a first round of PCR for *slβ* promoter with ccSLBp1s and ccSLBp2a, followed of a second round of nested PCR with ccSLAppGL3s and ccSLAppGL3a for *slβ* promoter wild type, and ccSLBppGL3s with ccSLBppEREs, ccSLBppEREa with ccSLBppGL3a for *slβ* promoter without ERE, in order to obtain desired fragments for Gibson Assembly reaction. The PCR mixture contained 1x PCR buffer, 0.5 µM of each primer, 0.2 mM dNTPs mix and 0.02 U/µL of Phusion® High-Fidelity DNA polymerase (New England BioLabs). PCR reactions contained 5 – 10 ng of template plasmid and were carried out under the following conditions: the initial denaturation of 30 s at 98°C, followed by 35 cycles of

10 s at 98°C, 30 s at 68°C, 40 s per kb at 72°C, followed by a final elongation of 5 min at 72°C. All amplified fragments were purified as described in 4.2.2.5. Fragments were assembled according to the one-step isothermal DNA assembly method (Gibson *et al.,* 2009) using NEB Gibson Assembly mix. The mixture was incubated at 50°C for 1 h in a thermocycler, and 5 μL of mixture were transformed as described on 4.2.2.13.

Name	Sequence	Target	Use
ccSLBp1s	TATGCCTATCTTCCTGCGCAG	Slβ promoter	Cloning
ccSLBp2a	CCTATTCAGCCAGATGTATGAC	Slβ promoter	Cloning
ccSLBppGL3s	tcgagatctgcgatctCAATCAGCCAAGCAATAAG	Slβ promoter	Gibson assembly
ccSLBppGL3a	ccggaatgccaagcttTGATTGTGGTTCTGTCCTAG	Slβ promoter	Gibson assembly
ccSLBppEREs	ctcgagtcgccatcgATGTGTCACTTTGTTTTTATGC	Slβ promoter	Gibson assembly
ccSLBppEREa	gtgacacatcgatggcgactcgagtcCTTACCAGCACTTACGTTC	Slβ promoter	Gibson assembly
pGL3s	AGATCGCAGATCTCGAGC	pGL3 vector	Gibson assembly
pGL3a	AAGCTTGGCATTCCGGTAC	pGL3 vector	Gibson assembly

Table 4. Oligonucleotides for reporter assay constructs.

4.2.2.11 Agarose gel extraction

In many cases, a simple column based purification of a PCR product is not possible or DNA fragments have to be separated after a restriction endonuclease digest. Hence, separation of the DNA on preparative agarose gels (see 4.2.2.15) was necessary. The desired bands were cut from the gel with a clean scalpel using the blue-light screen and an orange filter. Elution of the DNA from agarose plugs was carried out with the Wizard® SV Gel and PCR Clean-Up System from PROMEGA, according to the manufacturer's instructions.

4.2.2.12 Ligation

For all ligations, a 1:3 molar ratio of vector: insert was used, unless the insert was exceptionally big. In these cases, a 1:1 molar ratio was used. Ratios were calculated using Promega's BioMath. For compatible cohesive ends or blunt ends, vector and insert were directly ligated in 1x T4 DNA Buffer (PROMEGA), and 2.5 U/µl of T4 DNA ligase (PROMEGA). To ensure efficient ligation, the final DNA concentration should not exceed 10 ng/µl, e.g. to ligate 500 ng of vector + insert, a reaction volume of 50 µl is the minimum. Ligation was carried out over night at 4°C.

4.2.2.13 Transformation

For transformation of plasmid DNA to chemically competent *Escherichia coli* (*E. coli*) cells, an aliquot (50 µl) of the respective bacteria strain was thawed on ice. 50 ng of circular DNA were added to the cells and incubated for 30 min on ice. To transform a ligation product, the whole maximum 5 µL of ligation reaction was added. The heat-shock was performed for 45 s at 42°C in a Heat block. Tubes were put back on ice for 2 min to reduce damage to the *E. coli*

cells. Afterwards, 0.5 ml of pre-warmed (37°C) liquid LB medium (without antibiotics) was added and the cells were incubated at 37°C for 1 hour with agitation (300 rpm).

About 100 µl of the resulting bacterial culture were spread out on solid LB media with appropriate antibiotic(s) added. In general concentrations of 50 µg/ml amplicillin were used. Cells were grown at 37°C over night and plates were inspected after 12 – 16 h.

4.2.2.14 Colony PCR

An easy and quick way to screen for positive clones after transformation is performing colony PCR. Therefore, about 20 clones were picked from solid LBamp media plates after overnight incubation. Every colony was picked with a sterile pipette tip and resuspended directly in to PCR mix. The remaining cells at the tip were stroke out on fresh LB plates with appropriate antibiotic(s) added and incubated at 37°C over night. PCR reaction was carried out according 4.2.2.4 using the following program: Initial denaturation 95°C during 10 min, follow of 30 cycles of denaturation 95°C for 30 s, annealing 55°C for 30 s, extension 72°C for 30 s*, and Final extension of 10 min at 72°C.

* Extension time depended on the length of the expected product size. The Taq polymerase used here has an approximate elongation rate of 1 Kb/min. For a PCR product of e.g. 500 bp, an extension time of 30 s was applicable.

The PCR products were analyzed on 1% agarose gels containing 0.01% SYBR®Safe as described in 4.2.2.15. Positive clones were selected and about 4 clones were inoculated from the LB plate in 3 – 4 ml of liquid LB medium

supplemented with appropriate antibiotic(s). Liquid cultures were incubated at 37°C over night with agitation.

4.2.2.15 Agarose gel electrophoresis

In general, agarose gel electrophoresis was used to separate DNA according to size and for quality control of DNA isolations or PCR reactions. Agarose concentrations varied between 0.8 – 2%, dissolved in 1x TAE buffer supplemented with 0.01% SYBR®Safe. DNA was mixed with 6x sample buffer prior to loading the gel. As running buffer, 1x TAE was used. Gels were run at a constant voltage of 100 V until sufficiently separated. The separation process was monitored during and after the run by using blue-light emitting screens.

4.2.3 RNA extraction

Tissue samples of about 100 mg were mashed in a glass potter with a Teflon pistil and RNA extraction was performed according to Chomczynski and Sacchi (1987). RNA purity was assessed using spectrophotometrical measurements according 4.2.2.2.

4.2.4 Reverse transcription

Reverse transcription of 5 µg total RNA from each sample was performed using oligo-dT and MMLV reverse transcriptase (Promega) according to supplier's instructions.

4.2.5 RT-qPCR analyses

Specific primer pairs were designed for quantitative real-time PCR on cDNA template (RT-qPCR) and used for quantitative expression analyses of mRNA

levels of *slα*, *slβ* in pituitary (Table 5) or quantification of ChIP results for *slα* and *slβ* genes (Table 8). All amplicons were cloned and verified by sequencing. mRNA levels were quantified in cDNA by qPCR with BrilliantRII SYBRR Green, qPCR Master Mix (Agilent Technologies) according to supplier's instructions in a Mx3000 Real-Time Thermocycler. In a 40 cycles PCR reaction, each cycle consisted in 20 s at 94°C, 15 s at 55°C and 15 s at 72°C. Final heating from 55 to 95°C revealed melting curve of unique amplification product. All analyses were performed in duplicate. The expression level of each gene was normalized to β-actin expression as reference gene with exon-spanning primers to control for genomic DNA contamination since no DNAse treatment of total RNA was included. RT-qPCR assays were analyzed with 2(ΔΔCt) method (Livak and Schmittgen, 2001) via MxPro software (Stratagene) and expressed as Relative quantification to normalizer (Pfaffl, 2001). Statistical analyses of treated to control fish were performed using a Student's t test.

Gen	Acc. Number	Sequence	Amplicon size
β Actin	M24113	s 5´-ggacctgtatgccaacactg- 3´ a 5´ -gtcggcgtgaagtggtaaca-3´	In gDNA 383bp, cDNA 281bp
slα	GU434163	s 5´-ttctgactccgtccgtcttc- 3´ a 5´--gcaacaagctcagctgcctt- 3´	179bp
slβ	DQ021542	s 5´-tggattaatccactggtaga-3´ a 5´-tagacatcaatttgctcctg- 3´	91bp

Table 5. Oligonucleotides for qPCR quantification of both *sl* transcript levels.

4.2.6 Antibody development

An antibody specific for carp SLα was prepared against the synthetic oligopeptide NH_2-LIYLQTTLNRYDDAPK-COOH comprising amino acids 87-101 from the derived amino acid sequence of a carp SLα gene (GenBank accession no. ADE60529.2; Valenzuela *et al.,* 2015)). To increase immunogenicity, the peptide was crosslinked with *Concholepas concholepas* Hemocyanin (Biosonda) in 2.5% glutaraldehyde (v/v) in 0.5 M borate buffer, pH 9.0, overnight and dialyzed against 0.3 M NaCl at 4ºC. Equivalent to 200 mg were emulsified with 0.5 ml complete Freund's adjuvant (first injection) or incomplete Freund's adjuvant (subsequent injections). Intradermal injections were given every 3 weeks for a total of five injections. The rabbit was bled 2 weeks after the final injection.

4.2.7 Dot Blot and Western Blot

Total protein was extracted by homogenizing the pituitary gland in phosphate buffer pH 7.4 containing 1 mM EDTA and 1 mM EGTA (Invitrogen, California, USA) in the presence of PMSF (phenylmethylsulfonyl fluoride, Sigma Aldrich, St. Louis, MO, USA). The homogenate was precipitated at 20 ºC after adding five volumes acetone. After washing the precipitate with acetone, pellet was dissolved in sample buffer (25 mM Tris– HCl, pH 6.8, 5% glycerol, 1% SDS, 1.5% β-mercaptoethanol, 0.01% bromophenol blue) and analyzed by Western blot (12% separating gel). Proteins were transferred to PVDF membrane (Bio- Rad), and 1 h blocked at room temperature in PBS containing 5 % non-fat dried milk. Membranes were incubated at room temperature for 2 h with rabbit antisera (1:500) against SLα. After washing,

membranes were incubated for 1 h with a secondary antibody with biotin (Dako Universal Link) diluted in blocking solution, after washing incubated with streptavidin-peroxidase solution. Peroxidase activity was developed by adding H_2O_2 and diaminobenzidine as chromogenic substrate. Same procedure was used for dot blot.

4.2.8 Immunohistochemistry

Tissues were fixed in Bouin (7% picric acid saturated, 2.5% formalin 40%, and 0.5% acetic acid glacial 100%) for 24 h at room temperature, dehydrated, and embedded in Histosec (Merck, Darmstadt, Germany). Five micrometer thick sections were placed on gelatin coated slides and stored at room temperature. Tissue sections were dewaxed, gradually rehydrated, and treated with methanol (Caledon Lab., Georgetown, Ont., Canada) and hydrogen peroxide (Merck, Darmstadt, Germany) to block endogenous pseudoperoxidase activity. After washing sections were sequentially incubated with the primary antibody, using anti-carp SLα antibody (specific for SLα) and anti Salmon Coho SL (antibody detecting SL without distinguishing α or ß variant) kindly provided by Dr. Marianne Rand-Weaver at a 1:500 dilution in adjacent sections, for 4 h and finally with PAP complex (DAKO, Carpinteria, CA, USA). Peroxidase activity was developed with 0.1% 3,3`-diaminobenzidine (BRL, Rockville, USA) and 0.03% (w/v) hydrogen peroxide (Merck, Darmstadt, Germany).

4.2.9 Purification of Recombinant PIT-1

Complete coding sequence of *Cyprinus carpio pit1* based on Acc. Number AF132287 was amplified using following primers: PitPets

GGGTACCGGATCCATGACCTGCCAGGCCTTC and PitPeta CTGGTGCCGTCGACAGATCTACAGGCTGCGGT then cloned in to pETM between *Sal*I and *Bam*HI (NEB) restriction sites with an 6xHis Tag on C-Terminal and Flag epitope on N-terminal. The construct was transformed in to XL1-blue *E. coli* cells, purified and sequenced, and then transformed into Rosetta *E. coli* cells. Positive colony was inoculated in LB Medium and grown to OD_{600nm} = 0.5, then induced with 1 mM IPTG for 4 h at 30 ºC. Bacterial cells were pelleted and resuspended in Column Lysis Buffer, incubated with 1 mg/mL Lysozyme 30 min on ice, and then sonicated (10 times for 10 s at 50% on ice, with 10 s of rest on ice in between), centrifuged at 1000 xg 30 min to eliminate cell debris. Supernatant was incubated with 50% Ni-NTA beads (QiAgen) 1 mL per 4 mL of bacterial lysate for 1 hr at 4ºC. Later the flow of the column was opened and it was loaded with the different buffers from fraction 1-14 using Column Lysis Buffer (10 mM Imizadole), 15-30 Column Wash Buffer (20 mM Imidazole) and 31-41 Column Elution Buffer (250 mM Imidazole).

4.2.10 Denaturing Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Quantity and quality of protein preparations and cell lysates were analyzed using denaturing, discontinuous polyacrylamide gel electrophoresis (PAGE). SDS was used as charge shift molecule, which binds to proteins and imparts an even distribution of charge per unit mass. This allows fractionation of the proteins by size during electrophoresis. According to the proteins analyzed, separating gels of concentrations between 7.5 – 15% were used in combination with 5% stacking gels. To prepare the protein samples for the gel run, 1x Laemmli buffer was added and samples were heated for 5 min at 95°C. From prestained protein ladder PageRuler™ Plus (ThermoScientific)

4 µL were used as size marker. Gels were run in running buffer (0.25 M Tris/HCl pH6.8, 0.192 M glycine, 0.1% SDS) in and SDS-PAGE apparatus from BioRad.

4.2.11 Electromobility shift assay (EMSA)

4.2.11.1 Radioactive EMSA

Electromobility shift assays are a quick method to test for protein-DNA interactions *in vitro*. Small fragments of double-stranded DNA, containing putative binding sites for transcription factors, are prepared from synthetic oligonucleotides (40 pmol, 20 pmol/µL), one sense and the other antisense, annealed and labeled with [γ-^{32}P]-ATP (3000 Ci/mmol, 250 µCi, Perkin Elmer). The sense oligonucleotide (20 pmol, 20-mer) was first labeled at the 5'-end with polynucleotide kinase (PNK) (PROMEGA) and [γ-^{32}P]-ATP, discarding the excess of radioactive nucleotides by precipitation as described below: labeling was performed for 30 min at 37ºC in a total volume of 10 µL, was brought to a final volume of 50 µL adding 0.25 µL of glycogen (10 ng/µL, final concentration 50 ng/mL), 5 µL 3 M NaOAc pH 5.4 and supplemented with water, precipitated with three volumes of 100% ethanol, incubated 30 min at 20°C, centrifuged at 14.000 rpm (sigma centrifuge 2-16PK, rotor 12148 or equivalent) for 30 min at 4°C, the pellet was washed once with 70% ethanol and air-dried. The labeled sense oligonucleotide (20 pmol) was resuspended in 39 µL of water, 10 µL of 5x PNK buffer (PROMEGA) and 1 µL of antisense oligonucleotide (20 pmol), the mixture was heated in a water bath at 100°C for 10 min and cooled to room temperature during several hours standing on the bench. Competing unlabeled fragments were obtained by adding equimolar amounts of sense oligonucleotides (30 µL 100 µM) and antisense (30 µL, 100 µM), incubated in

mating buffer (70 mM Tris pH 7.6, 100 mM KCl, 10 mM $MgCl_2$ final), for 10 min at 100 °C, gradually cooled to room temperature and stored at -20°C. The binding reactions (Caccavelli *et al.,* 1998) were performed with 2 µg of nuclear protein extract, and 1 µL of DNA fragments labeled with [^{32}P] in 5 mM Hepes pH 7.8, 100 mM KCl, 0.5 mM DTT, 5% (v/v) glycerol. For competition experiments 200 molar excess of the DNA fragment containing the unlabeled binding site was added. The reactions were incubated at room temperature for 15 min and loaded on a prerun polyacrylamide gel (acrylamide: bisacrylamide 40:1) 4% (w/v) under native conditions. They were fractionated by electrophoresis at 4°C in 0.5x TBE buffer (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA) at 150 V, for 45 min, the gel was then dried and exposed on a radiosensitive screen, to then be revealed using a phosphoimager.

4.2.11.2 Fluorescent EMSA

The labeled single stranded DNA molecule was mixed with its complimentary non-labeled oligonucleotide in a ratio of 1:1.2 in annealing buffer. For that stock solutions were generated in concentrations of 10 µM and a total volume of 20 µL. The reaction was heated for 1 min at 95 °C and slowly cooled down to room temperature. 500 fmol of labeled (infrared or fluorescent dye), double-stranded DNA oligonucleotides were mixed with 250 – 1000 fmol of protein. In addition, 200 ng of poly(dI-dC) was added to every reaction to prevent unspecific binding of the protein to the target DNA. Reactions were carried out in EMSA buffer in a total volume of 10 µL and incubated for 10 min at 30 °C. Samples were resolved in orange G loading dye on 5% native PAGEs. Gels were pre-run in 0.4x TBE running buffer for 20 min at 80 V prior to application of the samples. Separation of free DNA and DNA-

protein complexes was performed for 1 h at 100 V and 4 °C.). Gels were imaged with an Syngene G-Box imaging system.

Name	Sequence	gene	T.F.
Site 1 s **a**	TCTCTGAGAATATTCACCAATCAC GTGATTGGTGAATATTCTCAGAGA	slα	Pit-1
Site 2 s **a**	AGATCCATAGATTCATAAACTCAC GTGAGTTTATGAATCTATGGATCT	slα	Pit-1
Site 3 s **a**	GAGTGCAATCAATAATGGCACCAA TTGGTGCCATTATTGATTGCACTC	slα	Pit-1
Site 4 s **a**	ATCGTGGCCAGCAACAGTTT AAACTGTTGCTGGCCACGAT	slα	Pit-1
Site 5 s **a**	GTATATAAGGTATGAATAACTAGG CCTAGTTATTCATACCTTATATAC	slβ	Pit-1
Site 6 s **a**	GGTTAATTCATAAATCCAGTAG CTACTGGATTTATGAATTAACC	slβ	Pit-1
Site 7 s **a**	GTTGAGATGATGCATAAAAACAAA TTTGTTTTTATGCATCATCTCAAC	slβ	Pit-1
ER1 s **a**	GACACATGTGGTCACAAATA TATTTGTGACCACATGTGTC	slβ	Estrogen Receptor
ERiso a **s**	CACTGACCTAATCAGGTCACATCATTATC GATAATGATGTGACCTGATTAGGTCAGT G	Carp Isotocin	Estrogen Receptor
ERh s **a**	GTAATCAGACTGGTCACATTCCCTC GAGGGAATGTGACCAGTCTGATTAC	Consensus	Estrogen Receptor
ER s **a**	GTAATCAGACTGGTCATGATGACCCATTC CCTC GAGGGAATGGGTCATCATGACCAGTCTG ATTAC	Consensus	Estrogen Receptor
Neg s **a**	GACTCGAGTCGACATCG CGATGTCGACTCGAGTC	Consensus	none

Table 6. Oligonucleotides used for transcription factor binding assays.

4.2.12 Transfection

Cell transfection was performed with Lipofectamine using standard procedures, as previously described (Páez-Pereda *et al.,* 2001). After plating the cells in 6-well plates, the cells were transfected in DMEM without FCS for 6 h using 5 µl Lipofectamine (Invitrogen) per well with 1.5 µg for each plasmid, in conjunction with a control (0.5 µg). Cells were then washed and left 24 h in DMEM supplemented with 10% (Fetal Calf Serum) FCS previously depleted of estrogen with charcoal.

4.2.13 Reporter assay

B6/GH3 rat pituitary cells were transfected as described in 4.2.12 in technical triplicates with plasmid containing the reporter gene *Firefly* luciferase generated according to the protocol 4.2.2.10 and SV40 driven *Renilla* luciferase control plasmid, was co-transfected to normalize the Firefly luciferase measurements. Thereby it is ensured that differences in expression are measured rather than differences in transfection efficiency. One control was used to determine the background of the reporter assay; non-transfected cells and cells co-transfected with pGEM-T-easy vector instead of the Firefly reporters. Cells were treated with 17β-estradiol at 24 h and harvested 48 h post transfection. The cell culture medium was aspirated and each well of the 12-well plate was washed twice in 1 ml of ice-cold PBS each. The 5x Passive Lysis buffer, a component of the Dual-Luciferase® Reporter Assay System kit, was diluted 1:5 in water. CHO cells were lysed by addition of 250 µL of 1x Passive Lysis Buffer per well and incubation for 15 min at room temperature on an orbital shaker. Reporter gene measurements were performed according to the manufacturer's instructions using a single-tube luminometer (Stratec Biomedical Systems, purchased via Biozym). Relative light units were

measured for *Firefly* and *Renilla* luciferase individually and ratios of *Firefly* RLU / *Renilla* RLU (F/R ratio) were calculated. Mean values of this ratio were calculated for the technical triplicates and the mean F/R ratio of the non-transfected control cells was subtracted from all transfected mean F/R ratios. All experiments have been repeated at least three times and the mean F/R ratios were averaged over all biological replicates.

4.2.14 Bisulfite sequencing

Bisulfite chemically deaminates non-methylated cytokines to uracil, leaving intact the methylated cytosines. For this, pituitary genomic DNA of male carp, treated with estrogen, which produces a differential effect on the expression of both SL genes (Valenzuela *et al.,* 2015), was extracted using phenol chloroform method and fractionated by sonication. This DNA was subjected to a bisulfite treatment using the Epitec kit (QiAgen) as follows: 1 ng – 2 µg of DNA were mixed with protection buffer and bisulfite mix, both provided in the kit. This mixture was denaturated at 95 °C for 5 min, then incubated for 25 min at 60°C, then re-denaturated for 5 min at 95 °C, incubated 85 min at 60 °C, denaturated 5 min at 95 °C and finally incubated 175 min at 60 °C. The treated DNA was purified with the columns provided in the kit, quantified and stored at -20 °C. This DNA was used as a template for a PCR reaction using primers flanking the promoter region of both SL (Table 7). The PCR products were cloned in pGEM-T Easy vector, transformed in to *E. coli* XL1-blue strain and 12 clones of each condition were isolated and sequenced (Macrogen, South Korea).

Name	Sequence	Target
bisSLA1s	GTTATTATTGATTGTATTTATGTATTTATTGG	slα
bisSLA2a	AATCCATAAATTCATAAACTCACC	Slα
bisSLA3s	GTATTTGTTAAAGGATTATATATGTAGGAG	slα
bisSLA4a	ATCAATTATTTATATTTAAAACAAACACAC	slα
bisSLA5s	TAAGGAAATTTGTTTATAGTAGTTTGG	slα
bisSLA6a	TCACAAACAAAATACATAACTCTTC	slα
bisSLA7s	TTTATAGTAGTTTGGAAGAAATGTATTAG	slα
bisSLA8a	TAAATAATAAAAATAAAATAATTCACTCAC	slα
bisSLB1s	TTAGTTTTAAATTAGATTTAAATGAAATATGG	slβ
bisSLB2a	CCTATTCAACCAAATATATAACTCTATACC	slβ
bisSLB3s	AGTATGTGTAATAATTAAATAAGTATTAGG	slβ
bisSLB4a	ATATATAACTCTATACCTATAATTTTCTTC	slβ
bisSLB5s	ATGGTTAATTGAATTATATTTTGTTAAGAG	slβ
bisSLB6a	AAACTCCTCTAACATATAATAAATAAACTC	slβ
bisSLB7s	GAATAGTTATATAGTATGTATATATATGTAGG	slβ
bisSLB8a	CACAAACAACAACTTAAACTCC	slβ

Table 7. Oligonucleotides for DNA methylation assays.

4.2.15 Chromatin immunoprecipitation (ChIP)

Prior to every ChIP assay in a new system, a mock experiment was performed to establish the optimal conditions for sonication. 40% output setting, 6 pulses of 15s, with 45s of rest on ice. The final size range of cross-linked DNA should be about 200 – 1000 bp. After reversal of cross-linking, the size of the sonicated DNA was checked by gel electrophoresis on 1,2% agarose gels as described in 4.2.2.15 (Figure S 3).

Pituitaries were cross-linked with Crosslinking Solution for 10 min at room temperature. The reactions were quenched with 125 mM glycine and incubated for 5 min at room temperature on a rocking platform. Then were washed twice in 10 ml of ice-cold PBS, Chromatin was extracted using ChromaFlash™ Chromatin Extraction Kit (Epigentek) according manufacturer instructions. Lysates were sonicated as described above at 4 °C Pituitaries lysates from 8 individuals weighing from 30 to 60 mg were pooled in two groups, control (n=4) and estrogen treated (n=4).

Cell debris was removed by centrifugation for 15 min at 13,000 xg and 4 °C. This step removes membranes, big proteins and precipitates as this particles sediment while chromatin and small proteins remain in the supernatant. The supernatant was transferred carefully to a new tube and 25 µL were removed as an 'input' sample. From Lysis buffer 475 µL was added to the input and it was stored at 4 °C. The remaining sheared chromatin was diluted with Lysis buffer, and pre-cleared incubating with prot. A/G Sepharose beads, 20 µL per 1 mL of chromatin, 4ºC 60 min, followed of a spin at 1200 xg 1 min at 4ºC. Supernatant was transferred to a fresh tube and double stranded DNA (dsDNA) concentration was determined as 4.2.2.2. Then diluted to a final

concentration of 2 mg/mL with 1x Lysis buffer 3 and 5% glycerol. Aliquots were stored at -80ºC.

Beads were prepared in a mixture of protein A/G-Sepharose in a ratio 1:2 due to different binding capacities, and stored at 50% (v/v) beads 50% (v/v) 20% ethanol. 100 µL of beads were precoated with 1x PBS, 20 µg/mL PolydIdC, 0.5% BSA, and then coated with 40 µg of each antibody, anti H3K4me3 (Abcam) and anti H3K9me3 (Abcam) for 30 min with constant rolling. Washed with 1x PBS and used for immunoprecipitation. ChIP samples were assembled in lubricated microcentrifuge tube, 25 µL of antibody-coated beads and 50 µL of pre-cleared chromatin extract, diluted to a final volume of 500 µL in Lysis buffer, and incubated 4ºC over night. Then Washed with RIPA Buffer five times, with TE buffer + 50 mM NaCl and eluted in Elution Buffer 10 min at 65ºC at 1400 rpm. Crosslink was reversed for four hours at 65ºC, incubated with 5 µg of RNase A and proteinase K final 0.4 µg/µL. DNA was purified with QIAquick PCR Purification Kit (QiAgen). 1 µL was used as template for qPCR reaction as described in 4.2.5 using primers of Table 8.

Name	Sequence	Target
ChIPSLA1s	CTGTACTGACTGTGCATCTT	slα
ChIPSLA2a	GCACCAAAAATAATGCAGGT	Slα
ChIPSLA3s	CTGGCATTCTCGTATTCTCA	slα
ChIPSLA4a	ACTCCTGCATATGTAGTCCT	slα
ChIPSLA5s	ACAGGCTTAAAGGTACACTG	slα
ChIPSLA6a	CGAGTGCTGTTTCATATCG	slα
ChIPSLA7s	ATCTTGTCATGTGTGTGTGT	slα
ChIPSLA8a	ACACGATCCAGAAGTTTCTC	slα
ChIPSLB1s	TCAGAAACGGCTGAATTTTG	slβ
ChIPSLB2a	GTAAGTGCTGGTAAGTCGAT	slβ
ChIPSLB3s	TGTACTTTGCGTTAGATGCT	slβ
ChIPSLB4a	ATTATTACACATGCTGCC	slβ
ChIPSLB5s	CGAGATGCTCAATTCATTCG	slβ
ChIPSLB6a	TCACGGAAATCTGTTGGAT	slβ
ChIPSLB7s	TGCAGGTTGAAAAATGTTCG	slβ
ChIPSLB8a	TTCATCTGGACAGTCAACAGT	slβ

Table 8. Oligonucleotide used for quantification of ChIP DNA

4.2.16 Statistical analyses

Assumptions of both normality and homogeneity of the variances were tested (SigmaPlot®11). The data were analyzed with Mann Whitney Test, for non parametric data, or with Student's T-test, for parametric data. $P<0.05$ was considered significant difference.

5. RESULTS

5.1. DIFFERENCES BETWEEN *slα* AND *slβ* GENES IN *Cyprinus carpio*

5.1.1 *sl* coding sequences

In order to understand the transcriptional regulation of *sl* genes, it was necessary to reveal gene sequences and the starting point were the coding sequences. Beside of this, coding sequence analyses unveil extra information about genetic divergences along evolution in teleost fish. Partial coding sequences previously published (Valenzuela *et al.,* 2015) were used as starting point, GenBank Acc. Num. GU434163 for *sl*α and DQ021542 for *sl*β.

To obtain *sl*α mRNA sequence a partial cDNA was amplified from male carp pituitary using specific primer designed against a known region with low identity score with *sl*β available sequence. Subsequently, primers designed for 5`-RACE Nested PCR amplified 390 bp and 39 bp products, respectively. Another cDNA was amplified from male carp pituitary using oligo-dT 3`-RACE and subsequently using specific primer designed against a known region with low identity score with *sl*β available sequence obtaining a product of 853 bp. The most complete *sl*α mRNA sequence obtained had 1402 bp, excluding the PolyA-tail, which was not identified in this experiment, and therefore could imply that full length *sl*α mRNA could be even longer. The *sl*α cDNA contains a 705 bp open reading frame encoding 234 amino acids, from the start codon (ATG), nucleotide position 33, to the termination codon (TAA), position 735. The cDNA sequence and predicted amino acids are depicted in Fig. 6. The predicted mature peptide contains 210 amino acids, and a signal peptide of 24 amino acids

s1α

```
AAACAACTGACCATTGACAAATCTGGGCAAAAATGAACAGAGTTAAAGTTCTGCAGGCCTGGCTGTGGTGTGTGTCTCTGTGGTCGTGCTGGCTTTCTCACGCC
                                  M  N  R  V  K  V  L  Q  A  W  L  W  C  V  S  L  W  S  C  W  L  S  H  A
ATTCCTCTGGACTGCAAAGACGAGACCGGATCACTCATCCAATGCACTTCAATCTCTCAAGAGAAACTTCTAGATCGCGTCATCCAACATGCAGAGCTCATCTA
 I  P  L  D  C  K  D  E  T  G  S  L  I  Q  C  T  S  I  S  Q  E  K  L  L  D  R  V  I  Q  H  A  E  L  I  Y
CCGCGTCTCTGAAGAGTCCTGTACTCTTTTTGAGGATATGTTTGTCCCGTACCCGCTGCATGCTTTGAGGAATCAGGGAGGAACCGTGTGGCACACCAAACCCT
  R  V  S  E  E  S  C  T  L  F  E  D  M  F  V  P  Y  P  L  H  A  L  R  N  Q  G  G  T  V  W  H  T  K  P
TCCCCATCCCAGGCTCCAAGACTGAGATTCAGCAGATTTCGGATAAATGGCTGCTTCACTCGGTCCTGATGCTGGTGCAGTCGTGGATGGAGCCCCTCATCTAT
F  P  I  P  G  S  K  T  E  I  Q  Q  I  S  D  K  W  L  L  H  S  V  L  M  L  V  Q  S  W  M  E  P  L  I  Y
CTGCAAACTACTCTTAATCGTTACGATGACGCCCCCGACGCTCTGCTGAACAAGACCAAGTGGGTGTCTGATAAACTGCTCAGCCTTGAGCAGGGGGTGGTGGT
 L  Q  T  T  L  N  R  Y  D  D  A  P  D  A  L  L  N  K  T  K  W  V  S  D  K  L  L  S  L  E  Q  G  V  V  V
CCTCATCCGCAAGATGCTGGATGAAGGGATTCTGACTCCGTCCGTCTTCGAGCACACGCTGACCCCATACGATGAACCGTCCCCAGAGGTGCTGGAGTCCATTC
  L  I  R  K  M  L  D  E  G  I  L  T  P  S  V  F  E  H  T  L  T  P  Y  D  E  P  S  P  E  V  L  E  S  I
TCAGGGACTACAACCTACTCATCTGCTTCAAGAAGGACGCCCACAAGATGGAGACCTTCCTCAAGCTGCTCAAGTGCCGTCAGAGCAACAAGCTCAGCTGCCTT
L  R  D  Y  N  L  L  I  C  F  K  K  D  A  H  K  M  E  T  F  L  K  L  L  K  C  R  Q  S  N  K  L  S  C  L
TCCTACTAAACCTGCATCTACTATCAAAAGAAAAAAGCTTTATTTTGAAGAATCAAACGTGTGCGTTGAATTGGCAGTAGGGGTAACTACAAAGTTAGCAAAGA
 S  Y  *
TTAAATGAGATCCCTTAGTGTGAACGCCGGAAATGGCTTTATTTAGTGTAACGTGAAGGGAACAGATGAAGATTCACACCGCTCTCACACACATCTGCACTTTT
GATGTCCTCCTTTCTCTTCCGAACTTGAGTTTGTATGCAATACAGACTTCACATAGTCTTTCAGTGTCTCTCCTGGTCTCTGGTGTGTTGGCAGGCATGTTTTT
TTTTTTTTTTTTTATAGTGGAGATGGTTTTACCTTTGTATTAGTTAGATGGAAAAGCTCTTGATACAGAAAATATGCTAATGAAGCCACAAAGCAGTGGCTGTG
TGTGTGTGTGTGCGTGCGTGCGTCTGTGGTGTGTGTGTGTGTGTGTAGGCGCATCACTGATTTGTTGACTCCTCCCTCTTTTTGTGACAGGCATCAATGTAT
CTTGCTGCTAGTAAGATACAGTCTTGTTTTTCACCTAATAAATACTGAAGTTGTGTTTTCTTGTTCAACATTTTGAAGTCAGGGTAAAGTATTTCAAATTTGGG
ATAAAATTTATTTCTAAGTTCTTCATAGTTAGAGCAGCTTGAGGAAGGTC
```

s1β

```
CACAATCAGAGAGTCCAACAAGACTTTTCTTTTGTTCTACGATAATGAAGAAAACTACAGGTCTACAGGTTTGTGTGGCGTTTGTGCTCTGCTCACTGCA
                                             M  K  K  T  T  G  L  Q  V  C  V  A  F  V  L  C  S  L  Q
GCCCGTGACTGGATCTACTGTTGACTGTCCAGATGAAGATACTGCAGGAACTTCCTGTACCATCTCACTGGAGAAGCTCCTGGAACGAGCCGTTCAACAT
  P  V  T  G  S  T  V  D  C  P  D  E  D  T  A  G  T  S  C  T  I  S  L  E  K  L  L  E  R  A  V  Q  H
GCAGAGCTTATTCACCACATGTCAGAGGAGTCCAAGCTGCTGTTTGACGAGATGCTCAATTCATTCGGTGTTGTGAATCTGCATATTTCCGAAGGGACCA
 A  E  L  I  H  H  M  S  E  E  S  K  L  L  F  D  E  M  L  N  S  F  G  V  V  N  L  H  I  S  E  G  T
TGTGTTCTCCTAAATCGGTGCCGGTTCCTATGTCTAAAACTGAAATCCAACAGATTTCCGACAAATGGCTCCTTCACTCAGTCCTGATTCTGGTCCAGTT
M  C  S  P  K  S  V  P  V  P  M  S  K  T  E  I  Q  Q  I  S  D  K  W  L  L  H  S  V  L  I  L  V  Q  F
CTGGATTAATCCACTGGTAGATGTGCAGGCATCTCTTATGAATTATGAGAATGTCCCAAGTGCCCTGCTTAACAGGAGCAAATTGATGCCTAGCAAAATA
  W  I  N  P  L  V  D  V  Q  A  S  L  M  N  Y  E  N  V  P  S  A  L  L  N  R  S  K  L  M  P  S  K  I
ATAAGCCTGGAGCAGGGTATACTGGTTCTTATTAGACAGATACTCGGTGAAGGTGGTTTGGTGTTGGAAGGAGCTGAAGACACATCCGATCGCTTTGTTT
 I  S  L  E  Q  G  I  L  V  L  I  R  Q  I  L  G  E  G  G  L  V  L  E  G  A  E  D  T  S  D  R  F  V
CTTCCGATATGCTTGAGACTGTGAGAAGAGACTACAGTGTGATCTACTGCTTCAGGAAAGATGCACACAAGATACAGACGTTCCTCAAACTGCTGAAATG
S  S  D  M  L  E  T  V  R  R  D  Y  S  V  I  Y  C  F  R  K  D  A  H  K  I  Q  T  F  L  K  L  L  K  C
CCGCCAGATTGATAAGGAGAATTGCTCCTTTTTCTAACTAAAAAAAAAATAATAATAATCTTCTGCCATATTATTTGGAGTGTTTCAGTGTTATTCTGCCA
  R  Q  I  D  K  E  N  C  S  F  F  *
AAGTTGGCTTTGCTTTTGCACAAATATTGAATTAAAAATAAAAAAATGTTTGCAAGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                       PolyA signal                         PolyA
```

Figure 6: *s1* transcript sequences. The cDNA sequences and deduced amino acid residues of *s1α* (Acc. Num: GU434163.1) and *s1β* (Acc. Num: DQ021542.1), including 5` and 3´UTR. Putative Polyadenilation signal and PolyA tail are highlighted on light blue.

(MKKTTGLQVCVAFVLCSLQPVTG). Despite putative polyadenylation site, AATAAA, was located at positions 1285–1290, it is unlikely that this is the right site, since it is very far from the end of the sequence where the PolyA-tail should be located. The 5`-untranslated region is 32 bp long, while the 3`-untranslated region has at least 665 bp. The calculated molecular mass of the mature peptide SLα is 24.239 Da with an isoelectric point of pI 5.61. The SLα mature protein contains six conserved cysteine residues, one less than other reported SL sequences (Zhu *et al.,* 2004).

In order to obtain *sl*β mRNA sequence using carp pituitary cDNAs obtained using specific primer designed against a known region with low identity score with *sl*α available sequence. Primers designed for 5`- and 3`-RACE amplified 492 bp and 478 bp products, respectively, using the same strategy which was used for *sl*α. The full-length *sl*β mRNA has 859 nucleotides, excluding the PolyA-tail (Fig. 6). The 5`-untranslated region is 44 nucleotides, while the 3`-untranslated region has 122 nucleotides, not including the PolyA-tail, but a polyadenylation signal, AATAAA, located at nucleotide position 837–842. The complete cDNA sequence and predicted amino acid sequence are depicted in Fig. 6, B. The *sl*β mRNA has a 693 bp open reading frame encoding 231 amino acids from the start codon (ATG), nucleotide position 45, to the termination codon (TAA), position 735, with a predicted signal peptide of 23 amino acids. The calculated molecular mass of the mature 207 amino acid SLβ peptide is 23.402 Da with an isoelectric point of pI 5.09. Mature SLβ has six out of the seven cysteines which are conserved with all other reported mature SL sequences such as zebrafish SLβ sequences (Zhu *et al.,* 2004).

SLα and SLβ share only 47.6% amino acid identity and 59% nucleotidic identity (Fig. 7), with cysteines residues conserved in position and number, except for fifth cysteine (in immature peptide, third in mature peptide), that is located at position 66 on SLα and position 87 in SLβ. Both derived protein sequences share highly conserved regions, depicted in black in Fig. 7 and also share the same number and position of putative N-glycosylation site, but differ on number and position of predicted O-glycosylation sites.

5.1.2 SL expression pattern

In order to analyze possible differential expression pattern of SL genes in carp pituitary we developed and prepared a SLα specific antibody, based on the herein predicted protein structure. In order to choose the best peptide for antibody generation, 3D structure prediction by SWISS-MODEL and hydrophobicity plots were used. Crossing this information with amino acidic sequence alignment between SLα and SLβ it was determined that NH_2-LIYLQTTLNRYDDAPK-COOH peptide was the best candidate to develop an antiserum specific against carp SLα.

Resulting antiserum was characterized first by dot blot (Fig. 8, A) observing no reaction in pre-immune serum, and specific reaction against pure peptide and against carp pituitary protein extract. Next step was Western blot analysis (Fig. 8, B) that showed two bands between 35 kDa and 25 kDa markers, this corresponded with predicted SLα and a putative glycosylated form, that was previously described in literature in other species (Rand-Weaver *et al.,* 1992).

Last stage was immunohistochemical detection of both SL variants (Fig. 8, C). For this adjacent sagittal sections of carp pituitary were used, one using

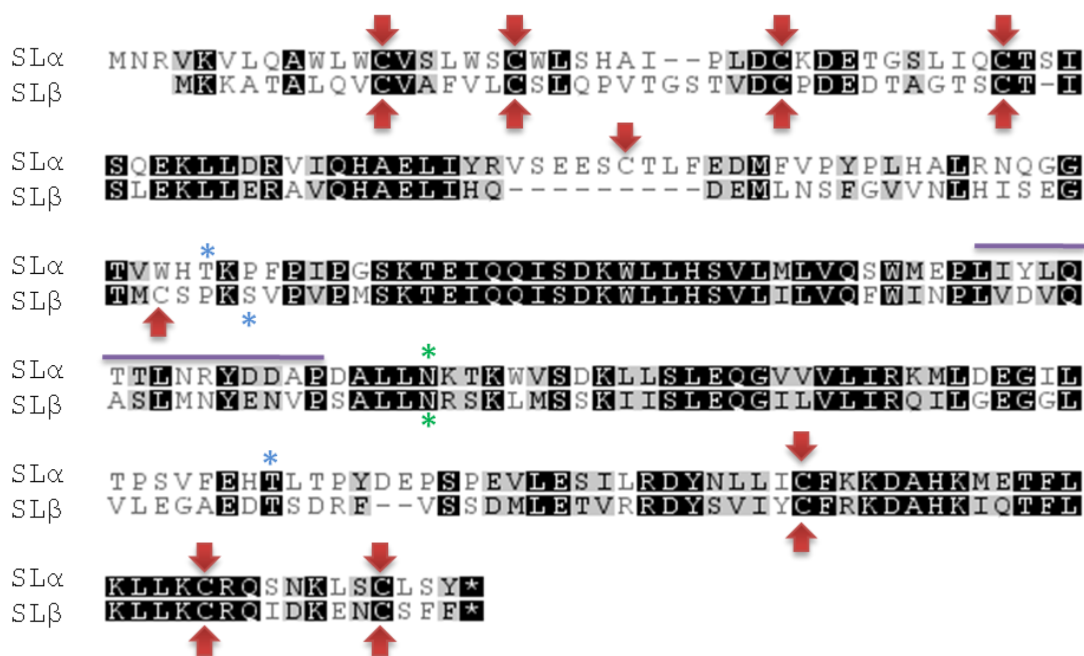


Figure 7: Alignment of SL amino acid sequences derived from two carp genes. SLα and SLβ alignment of derived aminoacid sequence. SLα 234 aa (Acc. Number: ADE60529.2) and SLβ 230 aa (Acc. Num: AAY45791.2) Similarity is represented in grey scale. Darkest residues represent strong similarity. Cysteine residues are remarked with red arrows, predicted N-glycosylation sites are marked with a green * or a blue * for the predicted O-glycosylation sites. Both sequences show 47,6% amino acid identity

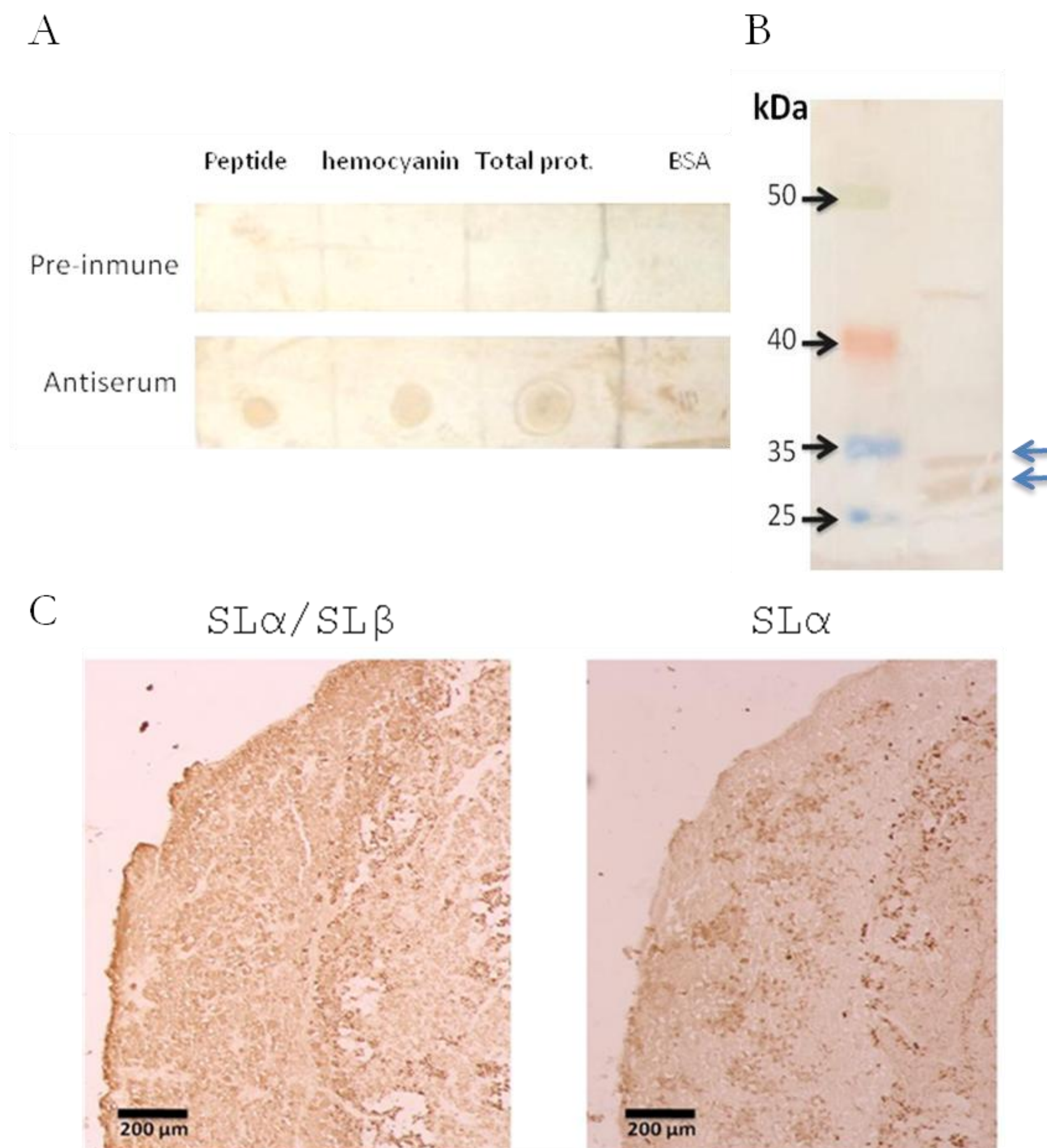


Figure 8: Differential spatial expression pattern of SL in carp pituitary. (A) Comparison of immunoreactivity between pre- and post-immunization serum by dot blot, using pure SL peptide, Hemocyanin (carrier), carp pituitary total protein extract and bovine serum albumin as negative control; (B) Western blot analysis in denaturing conditions of pituitary extracts of male *Cyprinus carpio* with SLα antisera 1:500, with corresponding molecular weight

marker. (C) SL immunodetection with anti-salmon-SL which does not discriminate α or β isoforms (*) and with specific anti-carp-SLα in adjacent sagittal sections of adult male carp pituitary. Images correspond to *pars intermedia*. * (Rand-Weaver *et al.,* 1993)

anti-Coho salmon-SL and another with anti-carp-SLα serum. In the first case as expected, immunoreaction corresponding to both SL proteins, was visible in cells exclusively located in neurointermedial lobe. In the second case the anti-SLα serum immunoreacted in *pars intermedia* which confirmed that the antibody reacted in that region specific for SL production, no reaction was detected in *rostral pars distalis* or *proximal pars distalis*. Furthermore and most important, clearly the anti-SLα displayed signal only in a subgroup of cells in comparison to the cells reactive with the anti-salmon-SL in *pars intermedia*. This suggest that SLα and SLβ proteins are produced in different sub-regions in carp pituitary neurointermedial lobe.

Next, to understand more about the regulation of the transcription of both *sl* genes, mRNA expression levels were quantified in pituitary under basal conditions. As shown in Fig. 9, expression levels of *sla* are significantly higher than *slβ* (approximately 2,5 times) in male carp pituitary. This added to the evidences in the literature that speak of a differential regulation in response to estrogen in male carp (Fig. 5) suggesting that both *sl* genes transcription are regulated differentially.

5.1.3 *sl* genomic sequences

Next step on characterization of regulatory differences between *sl*α and *sl*β, was reveal both genomic sequences, with focus on proximal promoter regions. Taking as starting point previously obtained coding sequences for inverse PCR, upstream genomic sequences were obtained and complemented with sequences *in silico* cloned from carp genome library (GenBank Acc. Num PRJNA73579).

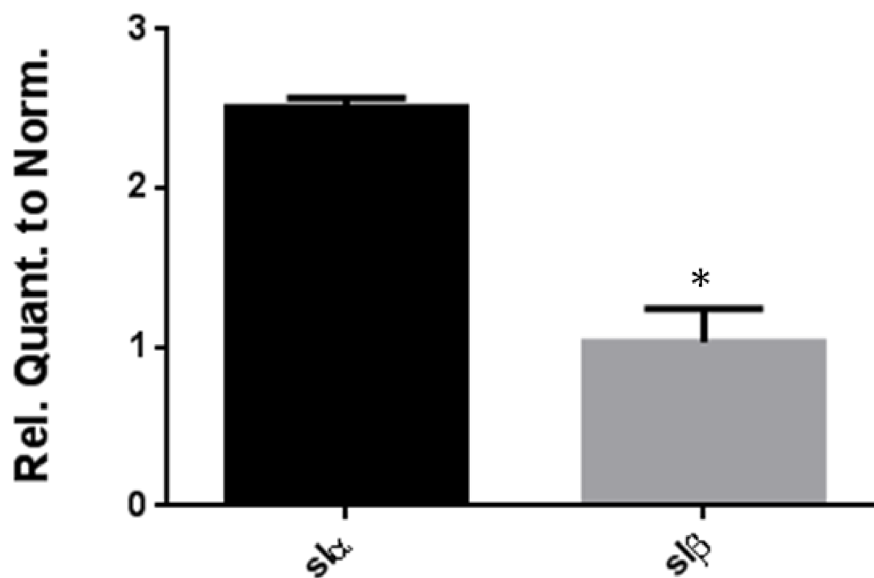


Figure 9: Differential expression of *sl* genes in adult carp. Graph depicts transcript levels of *slα* and *sl*β genes in pituitary of male carp normalized to β-actin expression (n=4).

The cDNA of the 5`-untranslated region obtained through 5`-RACE was aligned with genomic sequences to determine transcription start site TSS (Fig. 10). Sequence analysis of the *sl*α gene showed 5 exons, the sequences of which were identical to the corresponding cDNA. Exon 1 of 45 bp containing a short 5`-untranslated region and the ATG translation start codon, intron 1 242 bp, exon 2 193 bp, intron 2 711 bp, exon 3 114 bp, intron 3 2160 bp, exon4 180 bp, intron 4 69 bp, exon 5 at least 238 bp, containing TAA stop codon (the end of this exon was not determined due lack of enough genomic sequence) (Fig. 11). Comparative analyses revealed that the sizes of exons in *sl*α are very similar to those described in chum salmon and sea bream *sl*, although the introns present big length variation (Astola *et al.,* 2005). The *sl*α gene sequence analyzed was deposited in GenBank under accession number GU434163.3 (upgrading the previous version of the formerly available sequence).

On the other hand, *sl*β gene consists of 5 exons, exon 1 of 61 bp containing a short 5`-untranslated region and the ATG translation start codon, intron 1 281 bp, exon 2 184 bp, intron 2 140 bp, exon 3 115 bp, intron 3 105 bp, exon 4 178 bp, intron 4 110 bp and exon 5 321 bp, containing TAA stop codon and a putative polyadenylation site (Fig. 11). Size of exons are similar to those founded for *sl*α however introns showed important variations. The *sl*β gene sequence was uploaded to GenBank under accession number DQ021542.2 (upgrading the previous version of available sequence). The cDNA of the 5`-untranslated region was aligned with genomic sequences to determine transcription start site as is shown on Fig. 10. Further analyses with genomic sequences of both *sl* genes using a recently uploaded genomic assembly database (Acc. Num. GCF_000961615.1) determined that the *sl*α sequence obtained in this work, had 98% identity with an unplaced genomic

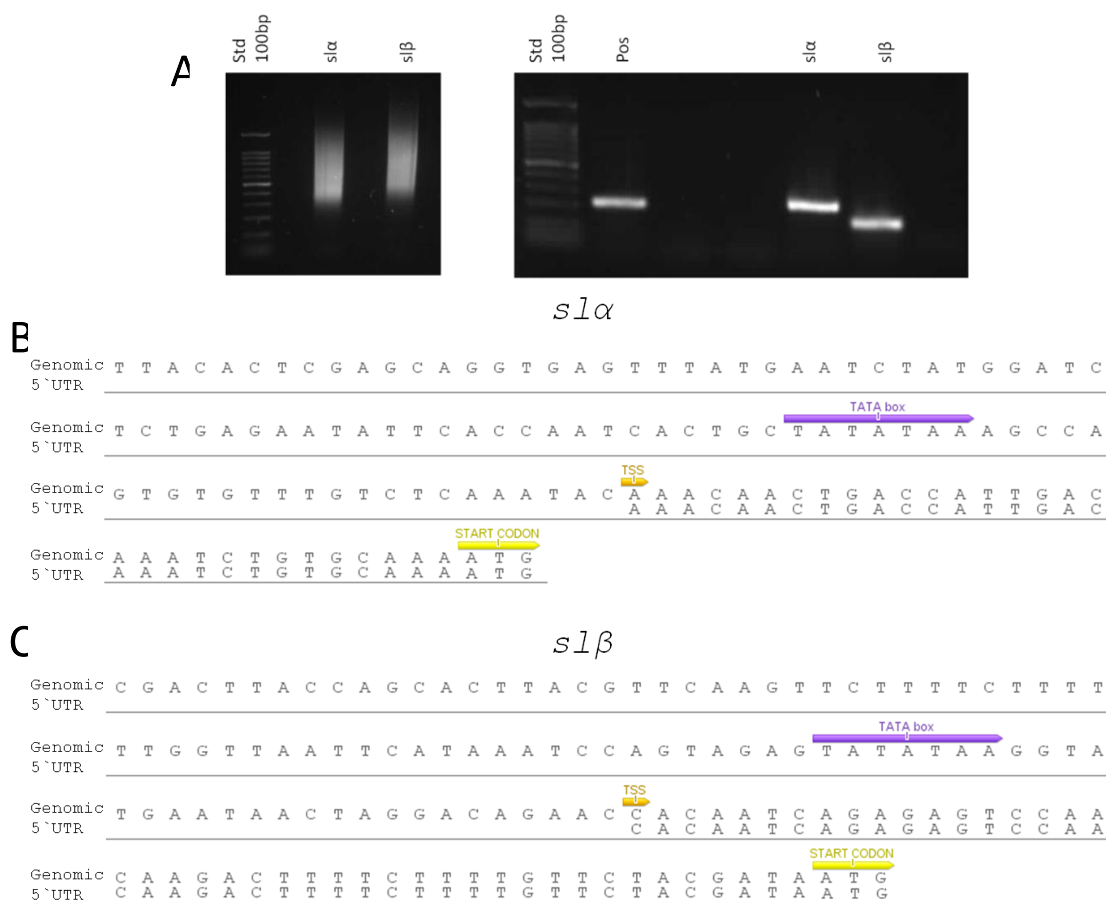


Figure 10: Identification of transcription start site of *slα* and *slβ* genes. (A) Identification of *slα* and *slβ* 5`-UTR sequence using 5`-RACE, images show electrophoresis of amplicons of consecutive nested PCRs on total RNA from carp pituitary extracts. 5`-UTR regions were amplified using a reverse primer located around the ATG/Met start codon and the main products were sequenced. Standard corresponds to 100 bp DNA Ladder (PROMEGA). (B and C) show alignment between genomic sequences of *slα* and *slβ* genes with corresponding 5`-UTR sequences, to determine transcription start site (TSS). TATA box are marked in purple, first codon is marked in yellow

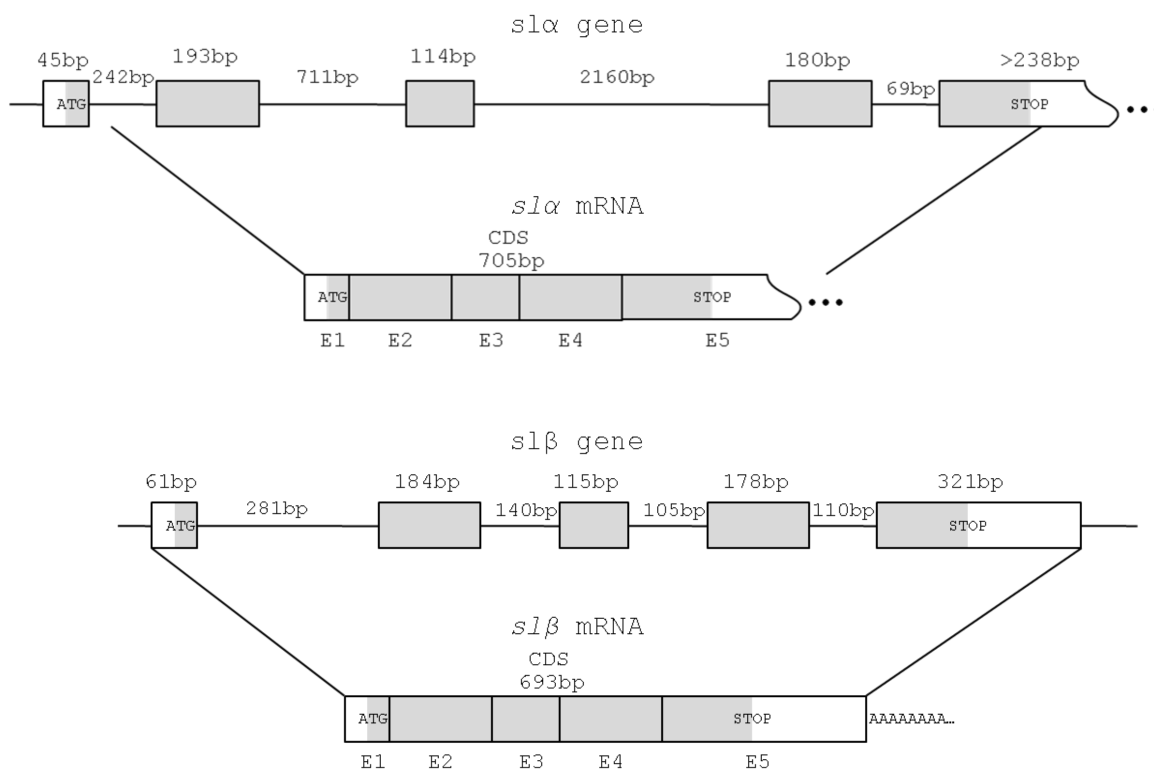


Figure 11: Genomic structure of *sl* genes. (A)Representation of *slα* (Acc. Num: GU434163.1) and *slβ* (Acc. Num: DQ021542.1) mRNA and genomic organization, based on alignments with sequences obtained from *Cyprinus carpio* Genome sequencing and assembly project, Acc. Num: PRJNA73579, scaffold117854.1 for slα and scaffold23927.1 for slβ, complemented with inverse PCR data. Coding region are painted in grey, 705 bp for *slα* and 693 bp for *slβ*. Exons are represented as boxes, ATG start codon was located in first exon in both genes.

scaffold (Acc. Num. NW_017544403.1) and 87% of identity with carp chromosome 44 (Acc. Num. NC_031740.1).

Also interestingly *sl*β genomic sequence obtained in this work had a 99% of identity with carp chromosome 1 (Acc. Num. NC_031697.1) suggesting both genes are located in different chromosomes.

5.2 SL REGULATORY ELEMENTS

In order to gain insight into the mechanisms of transcriptional control of the duplicated *sl* genes, the sequence of both *sl* genes proximal promoters was obtained. Analyses were focused on proximal promoter region (500 upstream transcription start site). For *sl*α gene, TATA box was identified on -23 to -30 position, and INR element with two mismatch respect to consensus sequence ($YYA_{+1}NWYY$) was located at the transcription start site. Taking into account the highest scores for consensus sequences obtained with three independent programs (Tess: http:// www. cbil.upenn.edu/tess/ and MatInspector: http://www.genomatix.de/cgi-bin/matinspector/ matinspector.pl) and EMBOSS 6.5.7 tfscan binding sites were predicted. Despite several putative transcription factor binding sites were found such as glucocorticoid receptor (GR), GATA-1, Sp1, c-Myb, this work focused on Pit-1 and ER binding sites. Four Pit-1 putative binding sites, first located at -42 to -50, second at -63 to -74 (a double site), third at -188 to -196 and fourth at -218 to -226. Hereafter named Pit-1 site 1, Pit-1 site 2, Pit-1 site 3, Pit-1 site 4, respectively. The consensus sequence of these sites closely matches the 8-bp motif T/ANCTNCAT described for fish Pit-1 (Ohkubo *et al.,* 1996). No

ER binding site was found. Just three CpG sites were identified at -86, -102 and -161 (Fig. 12).

In other hand *sl*β gene, TATA box was identified -24 to -30, and INR element with two mismatch respect to consensus sequence was located at the transcription start site. Three Pit-1 putative binding sites were identified, first located at -12 to -20, second at -41 to -52 (double site) and third at -122 to -134 (double site). From now named Pit-1 site 5, Pit-1 site 6, Pit-1 site 7, respectively. Interestingly, and ER putative binding site was identified at -101 to -105, but this site corresponds at only half (GGTCAnnn) of the consensus site (AGGTCAnnnTGACCT). Several other putative binding sites were founded, AP-1, NF-1, GR, Gata-2, Sp1, c-Myb, AP-2, GATA-2, including a Xenobiotic Response Element (XRE) at -229 to -234. Six CpG sites were identified, at -76, -93, -171, -232, -428 and -453 (Fig. 13).

Both promoters share only a 51,8% of identity, both have Pit-1 putative biding sites, but differs in position including, except for Pit-1 site 1 (in *sl*α) and Pit-1 site 6 (in *sl*β) that are located in both promoters 12 bp downstream TATA box (Fig. 14), this site is conserved amongst other species and even in other genes of the same superfamily as Growth hormone (Almuly *et al.,* 2005). This suggest that this particular Pit-1 site is relevant for the expression of *sl* genes. In other hand, ER putative binding site was found in *sl*β but not in *sl*α promoter, this could explain the differential regulation of *sl*β in response to estrogen previously reported (Valenzuela *et al.,* 2015).

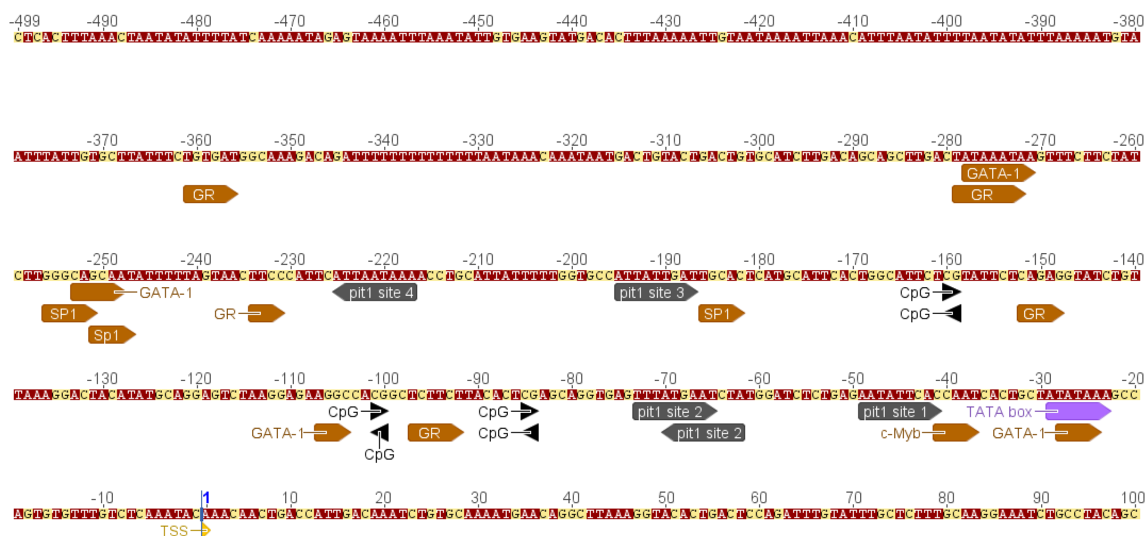


Figure 12: Carp *slα* proximal promoter sequence. 500 bp upstream transcription start site and 100 bp downstream are shown. TATA box is marked in purple. CpG dinucleotides are marked with black arrows. Transcription factors binding sites were predicted with EMBOSS 6.5.7 tfscan tool and marked with grey (Pit-1 binding sites) and brown arrows. A-T nucleotide are highlighted in red and G-C nucleotides with yellow.

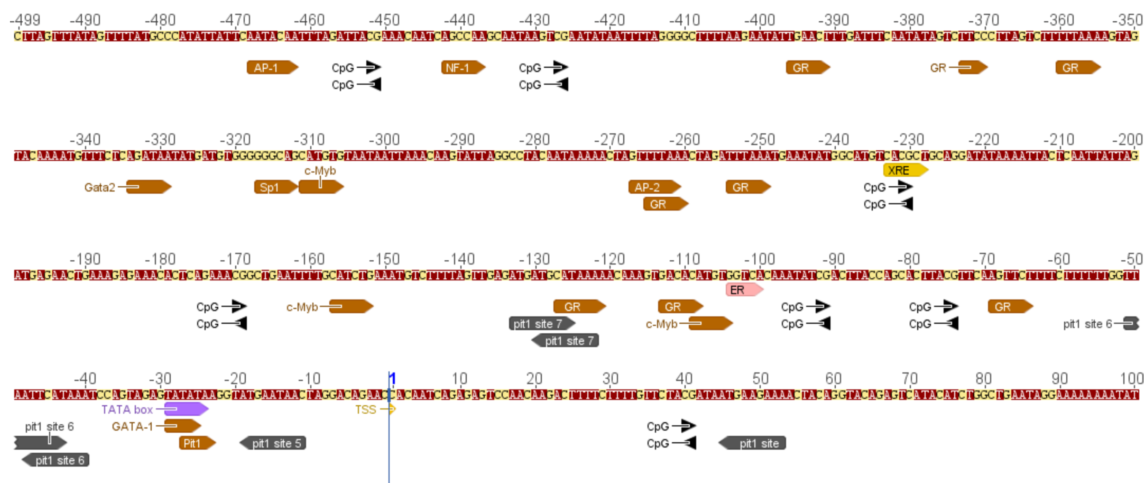


Figure 13: Carp SLβ proximal promoter sequence. 500 bp upstream transcription start site and 100 bp downstream are shown. TATA box is marked in purple. CpG dinucleotides are marked with black arrows. Transcription factors binding sites were predicted with EMBOSS 6.5.7 tfscan tool and marked with grey (Pit-1 binding sites), pink (Estrogen receptor binding site), yellow (xenobiotic response element) and brown arrows. A-T nucleotide are highlighted in red and G-C nucleotides with yellow.

Figure 14: Divergences on *sl* genes regulatory sequences. Sequence alignment between promoter regions (-500 to +100) from both *sl* genes. Conserved sites were shaded, start codons are depicted in yellow, transcription start site as TSS, TATA box in purple, CpG dinucleotides in black, putative Pit-1 binding site, Estrogen receptor binding site (ERE) and Xenobiotic response elements (XRE) in grey, pink and orange, respectively

5.2.1 Pit-1 and ER transcription factors Binding sites

To investigate the transcriptional regulation in response to estrogen of the two carp *sl* genes, the putative transcription factor binding sites found on isolated promoter sequences with focus on Pit-1 and ER binding sites. The binding potential of predicted sites was assessed using two strategies, purified recombinant Pit-1 and nuclear extracts from estrogen treated adult male carp. In the first case, carp Pit-1 coding sequence (Acc. Num. AF132287) was cloned into and His-Tag N-term vector (Fig. 15, A-C), expressed on *Escherichia coli* and purified using a nickel column and Imidazole gradient (Fig. 15, D). Purity of the recombinant Pit-1 was assessed by SDS-PAGE and immunodetection with anti His-Tag antibody (Fig. 15, E and F).

Complex formation between recombinant Pit-1 and double stranded oligonucleotides containing *sl* putative Pit-1 binding sites fluorescently labeled was assessed by mobility shift assays, using negative control with sequence unrelated with Pit-1 binding site (AdRa). Specific complex formation was observed for Pit-1 site 1, site 2, site 3, site 5, site 6 and site 7, no complex formation was observed for Pit-1 site 4 or negative control (Fig. 16, A).

The second strategy to assess binding potential of predicted Pit-1 and ER sites pituitary nuclear extracts from estrogen treated adult male carp and radiolabeled oligonucleotides were used. Out of seven potential Pit-1 sites tested (same as first strategy), specific complex formation was observed with Pit-1 site 1, site 2, site 3, site 5, site 6 and site 7 that disappeared in the presence of specific cold competitor, but not when incubated with unrelated oligonucleotide (Fig. 16, B). Consistently with previous results no specific complex formation was observed for Pit-1 site 4 (Fig. 16, B, third gel).

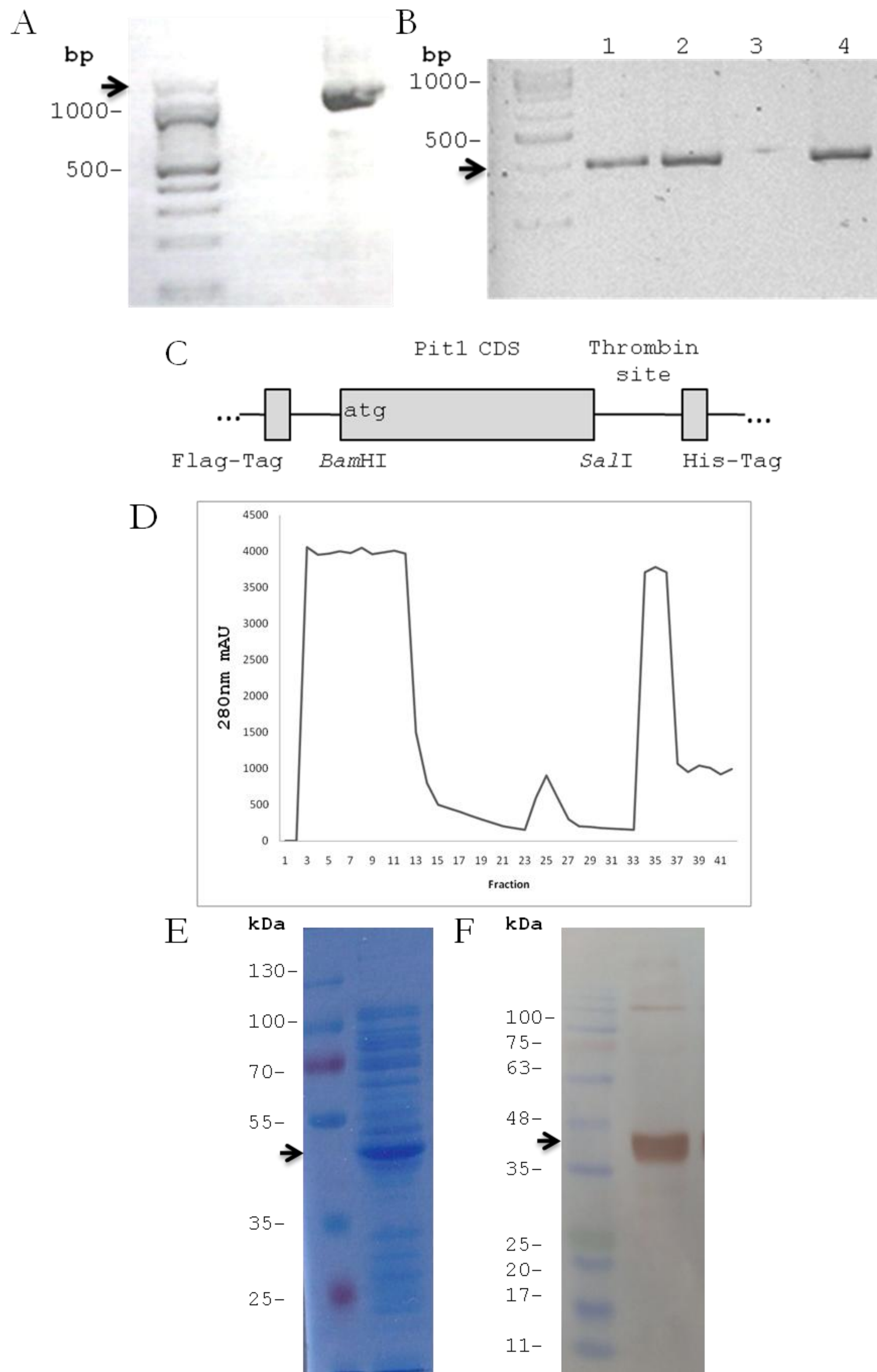

A
bp
1000-
500-
B
bp
1
2
3
4
1000-
500-
C
Pit1 CDS
Thrombin site
atg
Flag-Tag
BamHI
SalI
His-Tag
D
280nm mAU
Fraction
E
kDa
130-
100-
70-
55-
35-
25-
F
kDa
100-
75-
63-
48-
35-
25-
20-
17-
11-

Figure 15: Purification of recombinant PIT-1. (A) Amplification of complete coding sequence of carp *pit-1* based on Acc. Number AF132287, image shows an 1% agarose gel, Promega 100 bp DNA Ladder at left lane and, 1074 bp product of *Cyprinus carpio pit-1* amplification (marked with a black arrow). (B) Example of Colony PCR Screening, image shows an 1.2% agarose gel, Promega 100 bp DNA Ladder, amplification product of 349 bp of colonies containing plasmid with *pit-1* coding sequence inserted are marked with a black arrow and observed at lane 1, 2 and 4, meanwhile colony without *pit-1* plasmid is observed at lane 3. (C) Scheme of *pit-1* coding sequence insertion on pETM between *Sal*I and *Bam*HI restriction sites for production of recombinant PIT1 protein marked with 6x Histidine and Flag Tags. (D) elution profile of Ni-NTA His•Bind® Resins chromatography. Graph depict absorbance unit of each collected fraction. From fraction 1 - 14 using column lysis buffer (10 mM Imizadole), 15 - 30 column wash buffer (20 mM Imidazole) and 31 - 41 column elution buffer (250 mM Imidazole). Enriched recombinant PIT1 was collected from fraction 33 to 37. (E) SDS-PAGE of PIT1 containing chromatography fractions, stronger band at approx. 40 kDa correspond to recombinant His-Tagged PIT1, left lane correspond to PageRuler Plus Prestained Protein Ladder. (F) Immunodetection of recombinant PIT1 using anti His-Tag Antibody. Black arrow mark band at approx. 40 kDa that correspond to recombinant His-Tagged PIT1, left lane correspond to AccuRuler Prestained Protein Ladder.

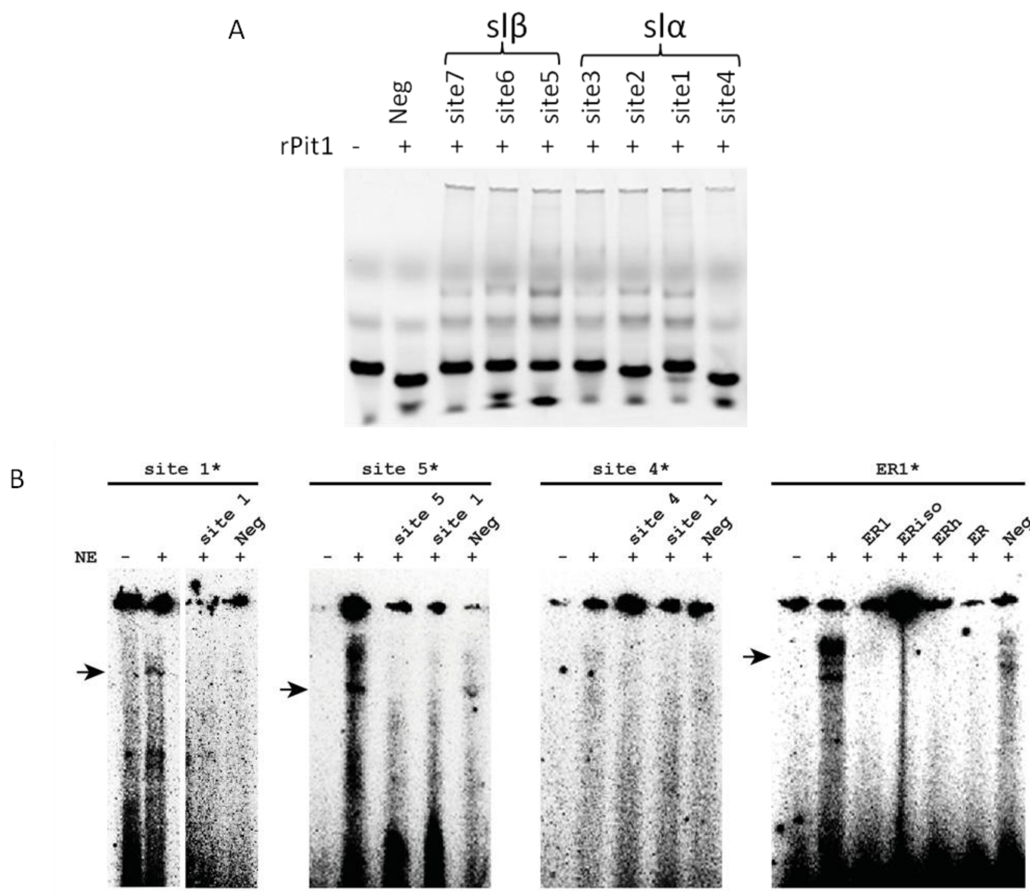


Figure 16: *In vitro* binding of nuclear factors to predicted Pit-1 and ER binding sites in *slα* and *slβ* promoters. With EMSA analyses specific binding was confirmed at site 1, 2 and 3 in *sla* and site 5, 6, and 7 on *slβ*, in (A) using recombinant Pit-1 (rPit1). In (B) an example of Pit-1 sites with EMSAs carried out with nuclear extract (NE), specific complex formation (arrow) confirming same sites than shown in (A) . Strong binding was observed on ccER1 from *slβ,* which was completely competed with general estrogen receptor consensus site (ER), as well as consensus half sit (Erh) and with ER derived from carp isotocin gene (ccERiso), which is under estrogenic control. Unrelated sequence (Neg) applied as non-specific competitor did not compete.

Also, Pit-1 differential binding affinities to binding sites in both *sl* promoters were assessed *in vitro* using serial dilutions of recombinant Pit-1 and Cy5 fluorescent labeled oligonucleotides. Strongest binding affinity was observed for Pit-1 site 3 in *sl*α promoter followed by site 5, 6 and 7 in *sl*β promoter. (Fig. 17). These results strongly suggest that site 1, site 2, site 3, site 5, site 6 and site 7 are indeed specific Pit-1 binding sites and that Pit-1 its relevant for both *sl* gene expression, as was described in literature for other members of GH and PRL superfamily (Li *et al.,* 1990).

Interestingly, when binding of ER putative site (identified at -101 to -105 at *sl*β promoter) was assessed with pituitary extracts, specific complex formation was observed that disappeared when the ER1 probe was incubated in the presence of different cold competitors, Pit-1 site 1 sequence, carp isotocin ER binding site, estrogen receptor consensus binding sites, half of estrogen receptor consensus binding site or the same ER1 sequence, but not of unrelated oligonucleotide (Fig. 16, B, fourth gel).

These evidences support the idea that this specific sequence indeed constitutes an ER binding site which might be involved on the previously described differential regulation of *sl*β in response to estrogen.

5.2.2 Influence of ER binding sites on estrogen responsiveness

To clarify the role of the ER binding site found in *sl*β promoter in the response to estrogen a dual-luciferase functional assay was performed. The dual-luciferase assay has been widely used in cell lines to determine rapidly but accurately the activity of a given promoter. In this case two constructs were assembled, based on pGL3 vector (Promega), one containing wild type

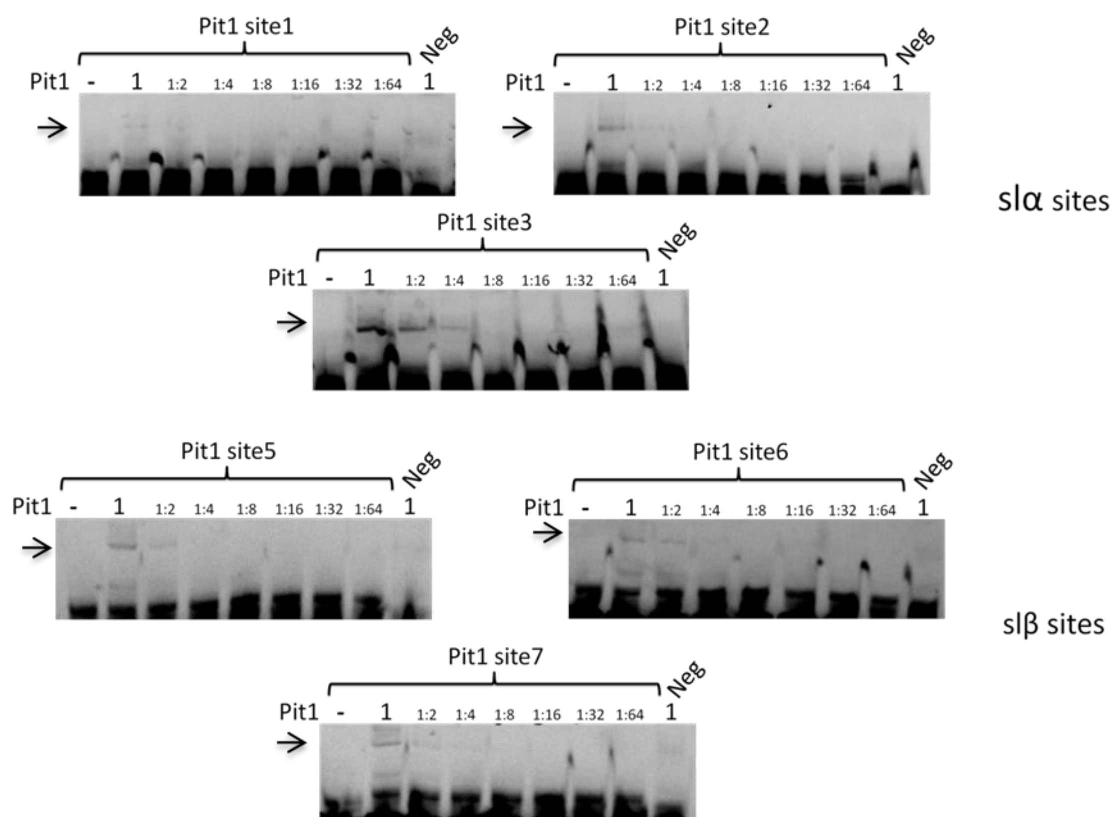


Figure 17: Pit-1 differential binding affinities to binding sites in *sl* promoters *in vitro*. Specific complex formation was detected by EMSA between double strand DNA oligonucleotides containing Pit-1 binding sites marked with Cy5 and serial dilutions of recombinant Pit-1 in order to compare binding affinities of Pit-1 sites in *sl* promoters. Strongest binding affinity was observed for site 3 in *slα* followed by site 5, 6 and 7 in *slβ*.

proximal *sl*β promoter and another without ER binding site, (Fig. 18, A) replaced for a sequences non related with ER binding site, but with same length (EREmu) in front of firefly luciferase coding sequence (Fig. 18, B). The assay was performed in rat pituitary cells GH3/BH6 cotransfected with a control *renilla* luciferase plasmid. Cells were treated with 17β-estrogen, and firefly luciferase was measured and normalized with renilla luciferase (Fig. 19, A). Clearly, wild type *sl*β promoter plasmid showed an statistically significant (p=0,003) increment on luciferase activity in response to estrogen, approximately 1.5 times, (Fig. 19, B) correlating perfectly with the *in vivo* increment on expression previously reported through RT-qPCR (Valenzuela *et al.,* 2015). However, cells transfected with EREmu plasmid showed no significant variation (p=0.229) in luciferase activity. This evidence point out that this particular ER binding site, located at -100 to -105 bp, is directly related with the differential expression of *sl*β in response to 17β-estrogen.

5.3. METHYLATION PROFILE OF *slα* AND *slβ* IN RESPONSE TO ESTROGEN

Given the relevance of DNA methylation as a significant regulator of gene expression (Ramsahoye *et al.,* 2000) was investigated whether this mechanism is associated with transcriptional regulation of both *sl* genes by estrogen in carp pituitary. To achieve this, gDNA from adult male carp treated with 17β-estradiol (n=4) or vehicle (n=4) was extracted from pituitary gland, treated with bisulfite, pooled and used as template for PCR reactions, in order to amplify promoter and coding sequences containing CpG sites to assess the state of methylation by sequencing (Fig. 20, left). Primer pairs and melting temperature for amplify regions containing CpG sites from bisulfite treated

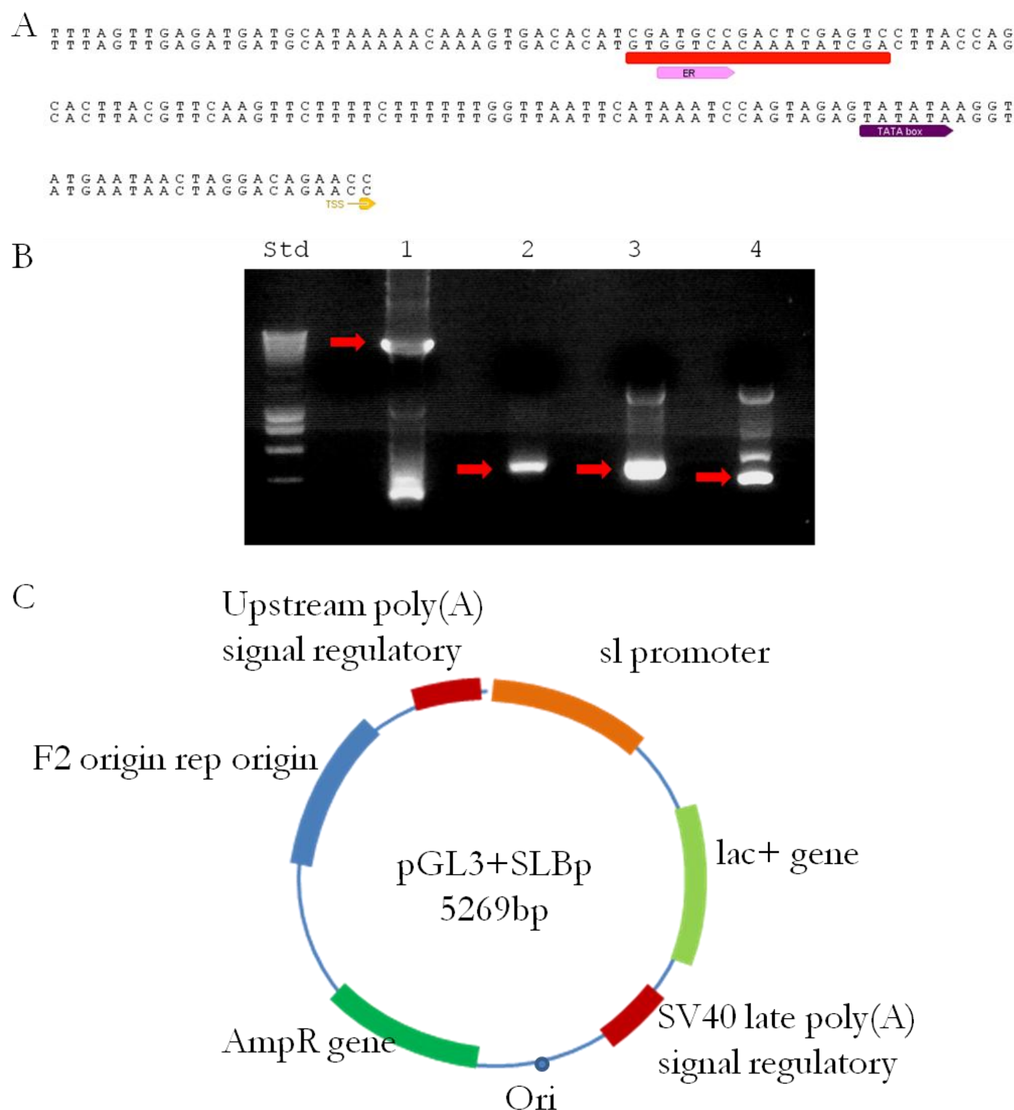


Figure 18: *sl* reporter assay construct design. (A) Promoter sequence inserted on pGL3 vector, containing *slβ* proximal promoter wild type (pGL3/SLB), and *slβ* proximal promoter without ER binding site (pGL3/SLB*EREmut*). Replaced sequence are underlined in red, predicted ERE half site its depicted in pink and transcription start site is marked with a yellow arrow. (B) shows electrophoretic separation of amplification products of PCR reaction for Gibson Assembly parts, pGL3 backbone (lane 1), *slβ* wild

type promoter (lane 2), and mutated ERE *s/*β promoter (lanes 3 and 4). (C) shows construct structure and key landmarks.

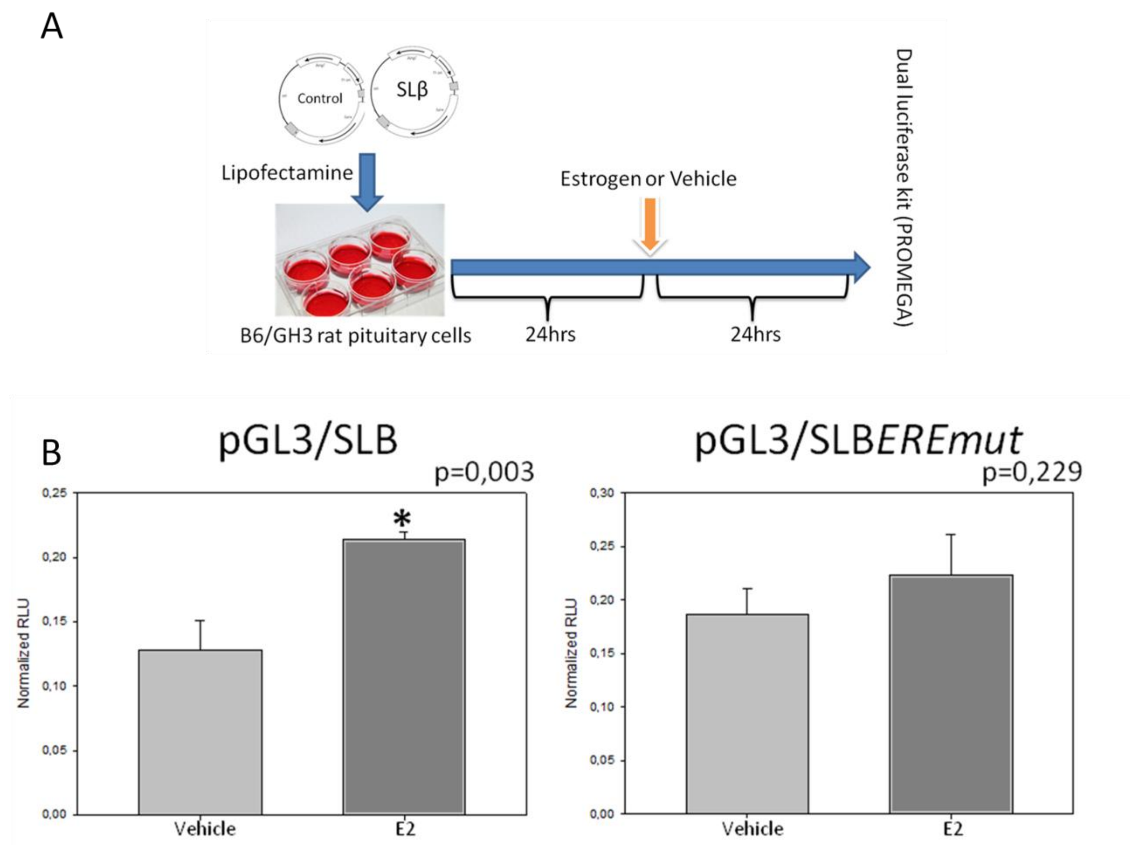


Figure 19: Analysis of *slβ* promoter activity using dual luciferase assay. (A) Scheme of *slβ* promoter reporter cell assay; (B) Graphs depict normalized relative light units corresponding to *firefly* luciferase activity in B6/GH3 rat pituitary cells transfected with a plasmid containing firefly luciferase under control of wild type *slβ* promoter (pGL3/SLB) and *slβ* promoter without ER binding site (pGL3/SLBEREmut) exposed to estrogen (E2) or vehicle, normalized to *renilla* luciferase activity, produced by a cotransfected control plasmid. Bar indicates standard deviation. (*) applies to significant difference. Student's t-test, P<0,05 was considered significant difference between controls and E2 treated cells.

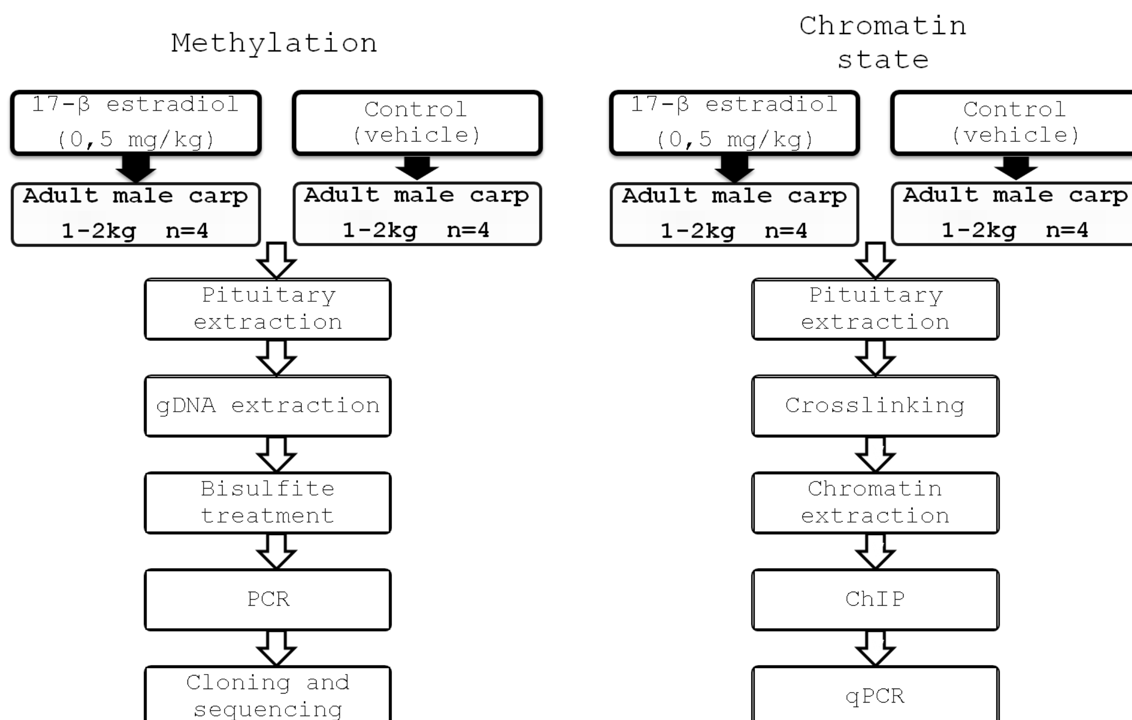


Figure 20: Scheme for epigenetic characterization of sl response to estrogen. Figure depict workflow to assess methylation levels (left), and chromatin compaction state (right) around *sl* proximal promoter. In both cases same experimental design was applied, taken from publication that first revealed differential expression of both *sl* genes in response to estrogen in male carp pituitary (Valenzuela *et al.,* 2015)

gDNA were standardized for *slα* (Fig. 21 A) and *slβ* (Fig. 22 A). Primers bisSLA7 and bisSLA6 were selected for *slα* promoter region, bisSLA2 bisSLA3 for coding region, amplifying 92 bp and 400 bp respectively (Fig. 21, B and C). PCR products were cloned in to pGEM®-T Easy vector (Promega) and transformed in to *E. coli* (XL1 blue strain). Colony PCR was carried out, using SP6 and T7 primers, to identify colonies that contain vector with desired insert (Fig. 21, D) for subsequent sequencing.

Primers bisSLB2 and bisSLB3 were selected for *slβ* promoter region, bisSLB6 bisSLB7 for coding region, amplifying 492 bp and 296 bp respectively (Fig. 22, B and C). Such as for *slα*, *slβ* PCR products were cloned and transformed, and subsequent colonies were screened through colony PCR (Fig. 22, D).

Regarding to *slα* gene two CpG sites were analyzed on promoter region, one located at -86, and second at -102 upstream of the transcription start site. Ten CpG sites were analyzed at coding region of *slα* gene, two first located at the first intron and the following eight at second exon (Fig. 23, C).

Methylation percentage of the two sites on promoter region was not affected by estrogen treatment, and remained around 20% (Fig. 23, A). 9 of 10 sites analyzed on *slα* coding region did not exhibit significant changes on methylation percentage, ranging from 62,5% (site 6 and 8 on control carp) to 100% (site 9 control carp), beside of site number 5 that showed an increment from 27% to 72,7% in estrogen treated respect to control carp (Fig. 23, B). Focusing on analyzed CpG sites, an important difference (more than 50%) of methylation percentage was observed between *slα* promoter and coding region, but no relevant effects of estrogen were observed on analyzed CpG sites. CpH methylation was also reviewed but no changes were observed in response to estrogen on promoter or coding sequence. Four CpG sites were

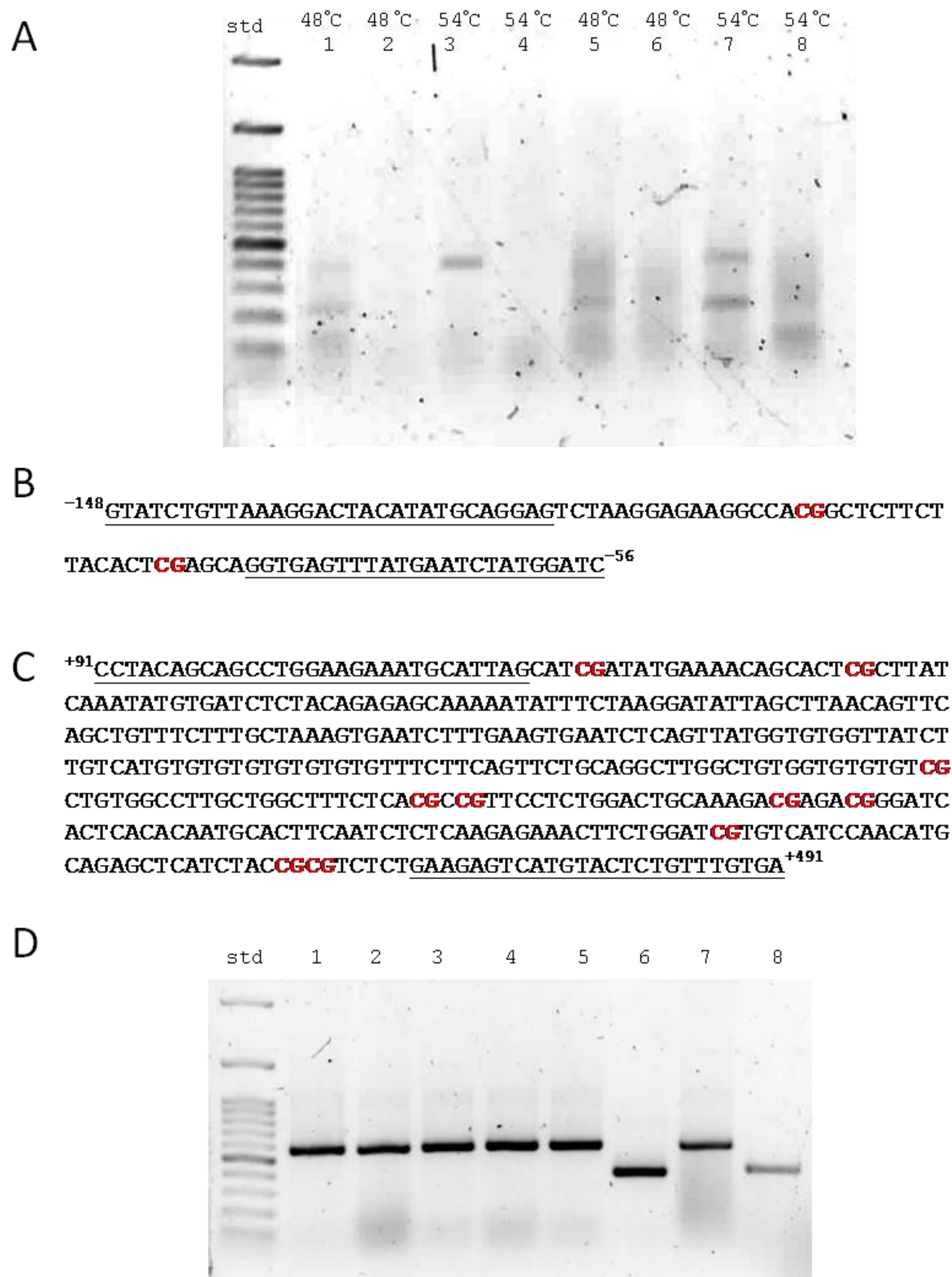


Figure 21: Methylation analyses procedure for ***slα*****.** (A) Example of standardization of primer pairs and melting temperature for amplify *slα* region containing CpG sites from bisulfite treated gDNA. Image depicts 1.2% agarose gel, AccuRuler 100 bp Plus DNA Ladder at left lane, followed by

amplification products using two different annealing temperatures, lane 1 and 3 product of bisSLA7 and bisSLA6, lane 2 and 4 product of bisSLA7 and bisSLA8, lane 5 and 7 product of bisSLA5 and bisSLA6, lane 6 and 8 product of bisSLA5 and bisSLA8. In this example primers are designed to amplify an CpG rich area in coding region of *slα* gene. (B) Sequence of selected region for methylation analyses in promoter of *slα*. CpG sites are in red, the sequence to which the primers chosen in the previous step (bisSLA7 and bisSLA6) are paired are underlined in black. Number indicate position respect to transcription start site. (C) Sequence of selected region for methylation analyses in coding sequence of *slα*. CpG sites are in red, the sequence to which the primers chosen in the previous step (bisSLA2 and bisSLA3) are paired are underlined in black. Number indicate position respect to transcription start site. (D) Example of colony screening PCR amplification products. Image depict 1.2% agarose gel, AccuRuler 100 bp Plus DNA Ladder at left lane, positive colonies, that contain vector pGEM T Easy with an insert of bisSLA7 and bisSLA6 primers amplicon, are observed in 1-5 and 7 lanes (557 bp using SP6 and T7 primers), negative colonies are observed in lane 6 and 8 (157 bp using SP6 and T7 primers).

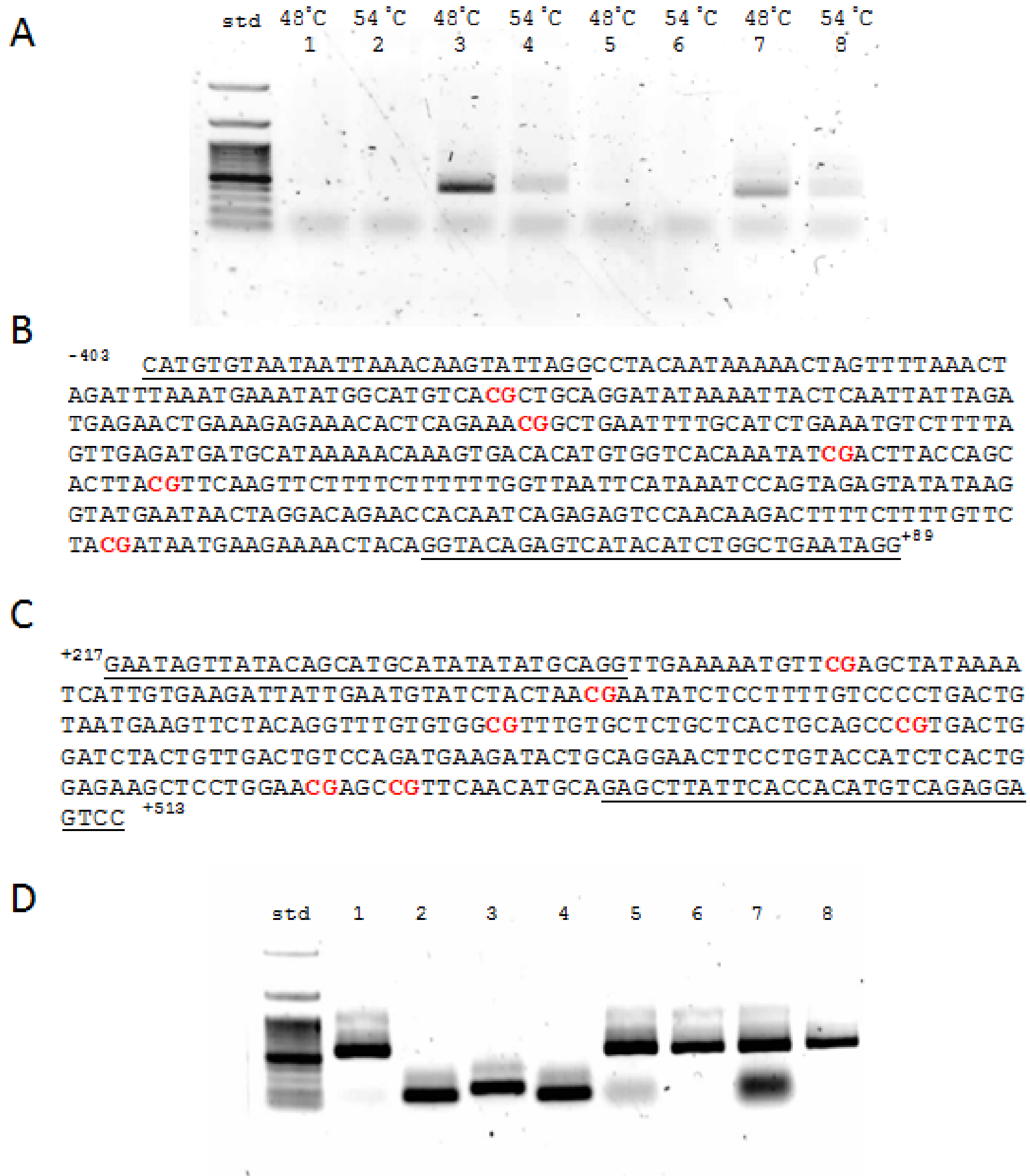


Figure 22: Methylation analyses procedure for *sl*β. (A) Example of standardization of primer pairs and melting temperature to amplify *sl*β region containing CpG sites from bisulfite treated gDNA. Image depict 1.2% agarose gel, AccuRuler 100bp Plus DNA Ladder at left lane, followed by amplification products using two different annealing temperatures, lane 1 and 2 products of bisSLA3 and bisSLA4, lane 3 and 4 product of bisSLB3 and bisSLB2, lane 5

and 6 product of bisSLA1 and bisSLA4, lane 7 and 8 product of bisSLB1 and bisSLB2. In this example primers are designed to amplify an CpG rich area in promoter region of *sl*β gene. (B) Sequence of selected region for methylation analyses in promoter of slβ. CpG sites are in red, the sequence to which the primers chosen in the previous step (bisSLB3 and bisSLB2) are paired are underlined in black. Number indicate position respect to transcription start site. (C) Sequence of selected region for methylation analyses in coding sequence of slβ. CpG sites are in red, the sequence to which the primers chosen in the previous step (bisSLB6 and bisSLB7) are paired are underlined in black. Number indicate position respect to transcription start site. (D) Example of colony screening PCR amplification products. Image depict 1.2% agarose gel, AccuRuler 100 bp Plus DNA Ladder at left lane, positive colonies, that contain vector pGEM T Easy with an insert of bisSLB2 and bisSLB3 primers amplicon, are observed in 1, 5, 6, 7 and 8 lanes (560 bp using SP6 and T7 primers), negative colonies are observed in lane 2, 3 and 4 (157 bp using SP6 and T7 primers).

analyzed on *sl*β promoter region, at -76, -93 (close to ER binding site), -171 and -232. On the other hand seven CpG sites were analyzed on coding region, one located at first exon, two located at first intron, and last 4 located at second exon (Fig. 24, C). On the promoter CpG sites no clear differential methylation was apparent in response to estrogen in these experimental conditions, site 4 was methylated in approximately 90%, meanwhile sites 3 and 2 around 60% and site 1 just around 30% (Fig. 24, A). In coding region all besides site 7, seemed to have alterations on methylation levels less than 20% in response to estrogen exposure. On the other hand, site 7 increased methylation 37.5% on estrogen treated carp (Fig. 24, B). These findings inform for the first time about methylation pattern on *sl*β in response to estrogen.

Comparison of methylation pattern of both *sl* genes clearly substantiated that the increased general methylation levels on *sl*β promoter region respect to *sl*α promoter in both control and treated carp, correlated with the superior expression levels of *sl*α observed in basal conditions (Fig. S 4).

5.4. *slα* AND *slβ* CHROMATIN STRUCTURE *IN VIVO* IN RESPONSE TO ESTROGEN

In order to evaluate the possible differences of chromatin compaction state in promoter of both *sl* genes in response to estrogen chromatin immunoprecipitation (ChIP) assays and subsequent quantification by qPCR were carried out. Pituitary from adult male carp treated with 17β-estradiol (n=4) or vehicle (n=4) were extracted, pooled, crosslinked and

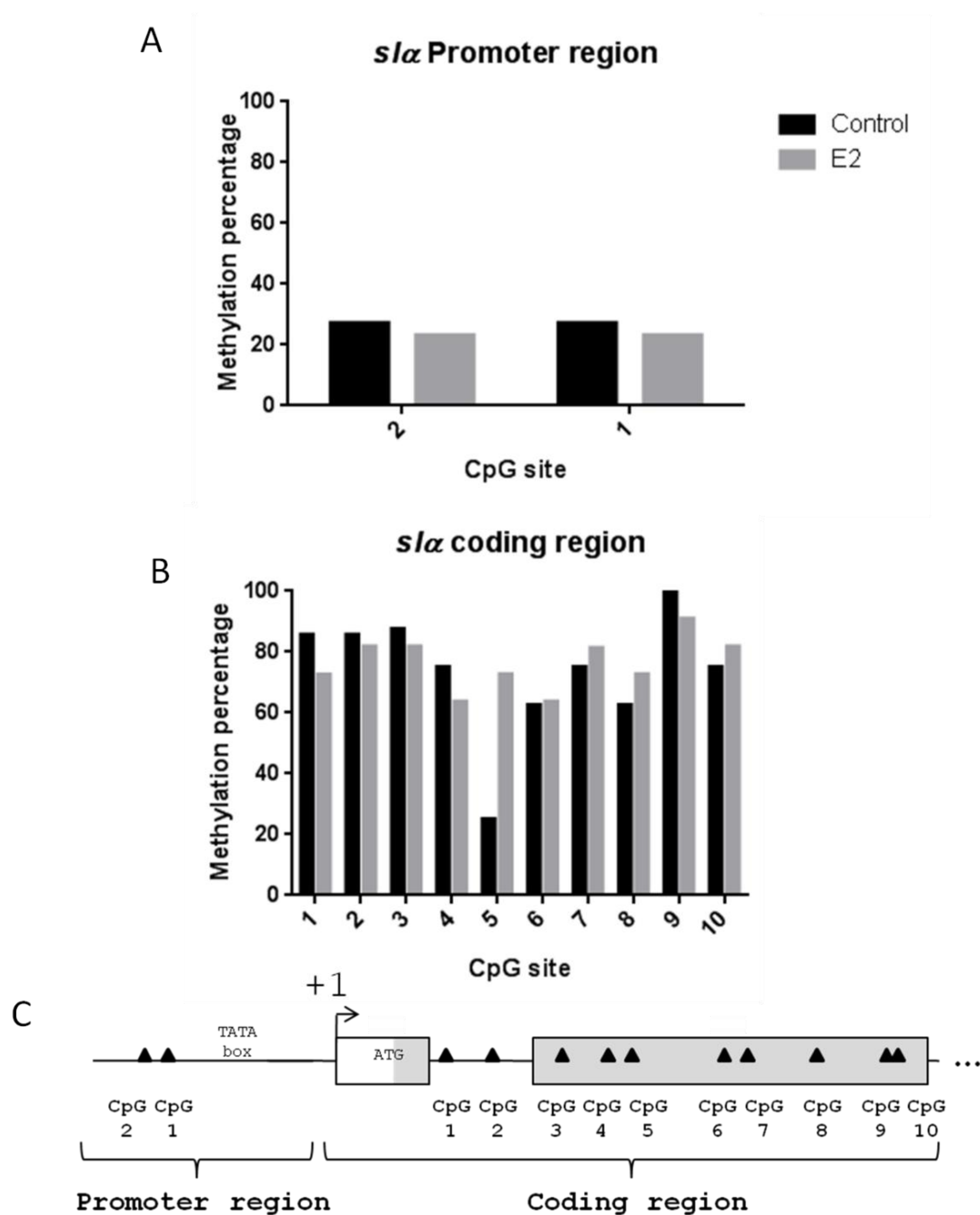


Figure 23: Methylation levels on *slα* gene. Graph depict methylation percentage of CpG sites in slα gene pituitary of adult male carp, at (A) promoter and(B) coding regions. Black bars represent methylation in control samples (pool of 4 individuals, 12 clones sequenced per condition), grey bars

represent methylation in estrogen treated samples (pool of 4 individuals). (C) Schematic representation of analyzed area in *slα* gene, highlighting key landmarks such as Exons, transcription start site, TATA box. CpG locations are marked with a black triangle.

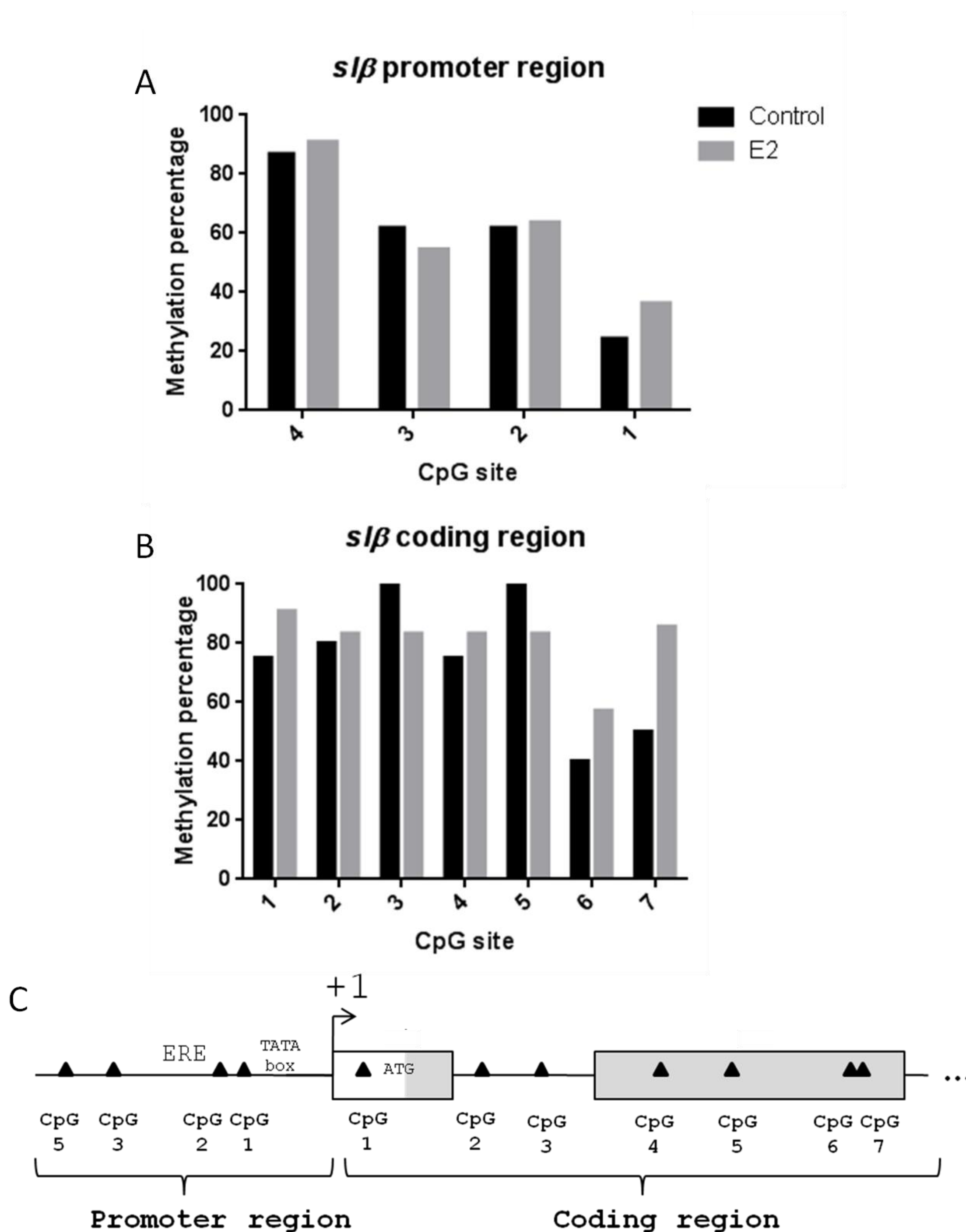


Figure 24: Methylation levels of *slβ* gene. Graph depict methylation percentage of CpG sites in *slβ* gene in pituitary of adult male carp in (A) promoter and (B) coding regions; black bars represent methylation in control samples (pool of 4 individuals), grey bars represent methylation in estrogen

treated samples (pool of 4 individuals, 12 clones sequenced per condition). (C) Schematic representation of analyzed area in *sl*β gene, showing key landmarks such as exons, transcription start site, TATA box. CpG locations are marked with a black triangle.

immunoprecipitated with antibodies against, Histone 3 trimethylated at lysine 4 H3K4me3 (active chromatin) and Histone 3 trimethylated at lysine 9, H3K9me3 (inactive chromatin) (Fig. 20, B). First step was design and standardization of qPCR primers to establish optimal conditions for the ChIP assay. Two primer pairs were design for promoter and coding region each, for both *sl* genes (Fig. 25, A), and specificity was monitored resolving the PCR products on agarose gel, and following primer pairs were selected: ChIPSLA1s-ChIPSLA2a, for *sl*α promoter region, ChIPSLA5s-ChIPSLA6a for *sl*α coding region, ChIPSLB1s-ChIPSLB2a, for *sl*β promoter region, ChIPSLB5s-ChIPSLB6a for *sl*β coding region (Fig. 25, B).

ChIP qPCR target amplifications were normalized for amplification of input DNA, and expressed as percentage of input. These results showed that H3K9me3, marker for heterochromatin, that seems to be less enriched on *sl*β promoter region in response to estrogen than in *sl*α promoter. However, H3K4me3, marker of euchromatin showed no important differences. Suggesting that *sl*β promoter region chromatin its less compacted than *sl*α promoter in response to estrogen treatment (Fig. 26). This correlates with expression data previously reported (Valenzuela *et al.,* 2015), and reporter assay findings (Fig. 19).

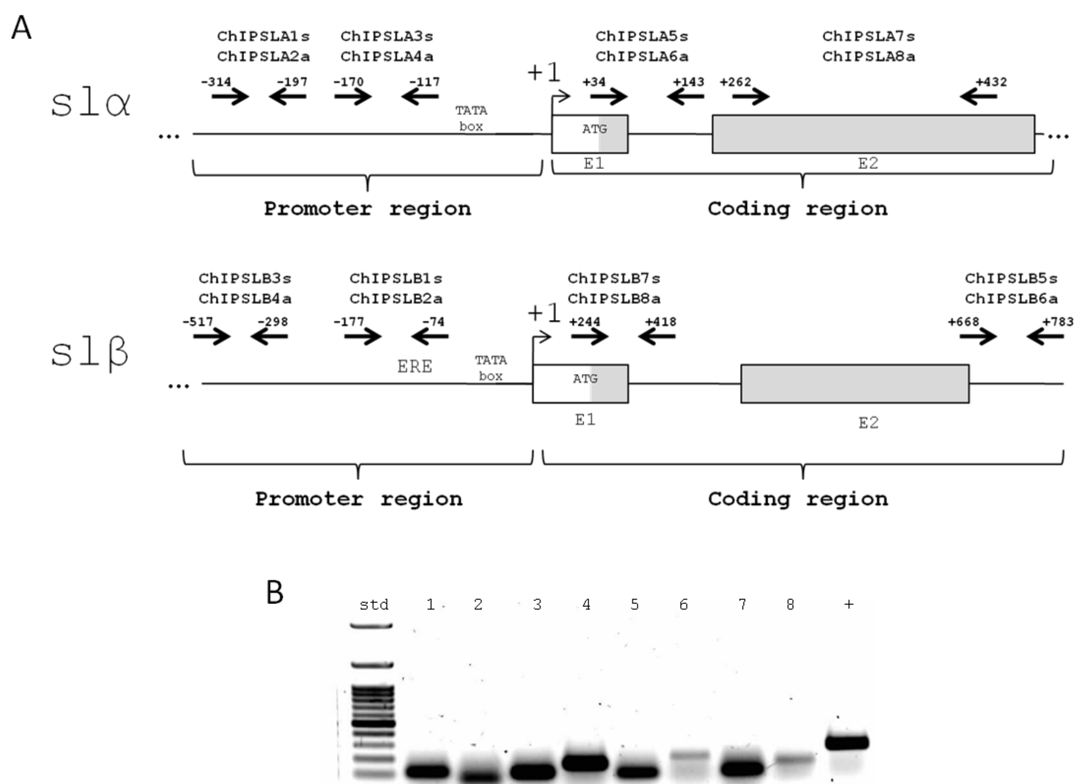


Figure 25: Primer design and standardization for qPCR for *sl* ChIP. (A) Scheme of localization of primers designed to quantify immunoprecipitated DNA from both *sl* genes in promoter and coding regions. (B) qPCR products separated on 1,2% agarose gel. AccuRuler 100 bp Plus DNA Ladder - Maestrogen, Lane 1: amplification product of chipSLA1s and chipSLA2a primers, lane 2: chipSLA3s and chipSLA4a,both directed against slα promoter region, lane 3: chipSLA5s and chipSLA6a, lane 4 of chipSLA7s and chipSLA8a, both directed against slα coding region, lane 5: chipSLB1s and chipSLB2a, lane 6: chipSLB3s and chipSLB4a, both directed against *sl*β promoter region, lane 7: chipSLB5s and chipSLB6a, lane 8: chipSLB7s and chipSLB8a, both directed against *sl*β coding region.

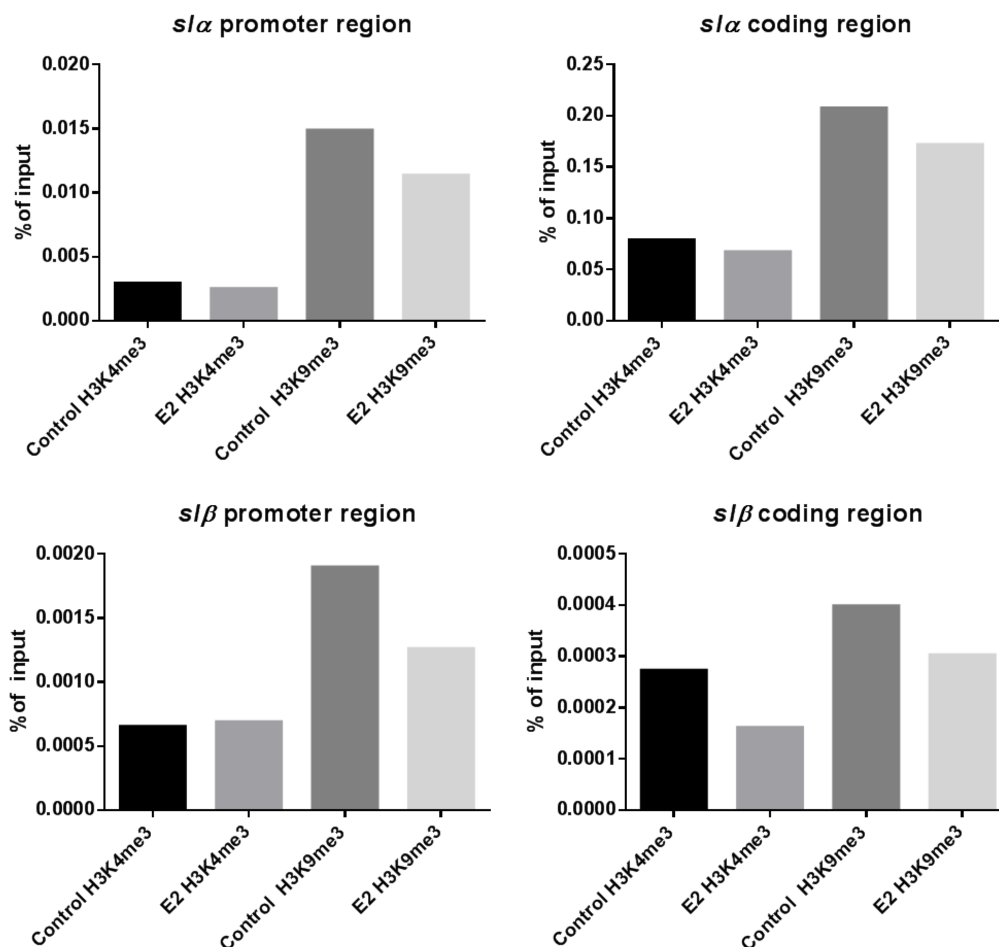


Figure 26: Chromatin compaction state around *sl* promoters. Graph depicts enrichment of H3K4me3 (active chromatin) and H3K9me3 (inactive chromatin) modified histone at coding and promoter region of *sl*α and *sl*β gene showed % of input DNA, quantified through qPCR measurements of the immunoprecipitated DNA of estrogen treated (n=4) and control (=4) carp.

6. DISCUSSION

SL is a fish exclusive pituitary hormone which belongs to the GH hormone family. SLα has been identified in many fishes; however, SLβ is characterized in a limited number of species which spend all or part of their lifecycle in freshwater (Zhu *et al.,* 2004). Expression of both *sl* genes was reported in pituitary of *Cyprinus carpio*, only *sl*β but not *sl*α responded with increased mRNA levels in pituitary of male adult carp to 17β-estrogen treatment respect to control suggesting a subfunctionalization or a subneofunctionalization. However, the mechanisms involved in this phenomenon are still unclear, so the objective of this study was to deepen the knowledge of this mechanism. In addition, this study contributes very necessary information for strategies when setting up biomarkers for monitoring of environmental changes with gene expression analyses, since the complexity of genomes of model organisms has to be considered and is of pivotal importance for development of quantitative bioassays.

In this work both *sl*α and *sl*β coding and regulatory sequences were characterized in the common carp, a worldwide distributed freshwater fish. SL is supposed to have existed prior to the last whole genome duplication event, from which the paralogues of *sl*α and *sl*β were derived (Zhu *et al.,* 2004; Benedet *et al.,* 2008) which is in accordance with the point from the newly recent report (Ocampo Daza and Larhammar, 2018). This fish-specific genome duplication was proposed after the finding that fish had seven Hox gene clusters, almost twice as many as primates and murines (Amores *et al.,* 1998; Prince and Pickett, 2002; Seoighe, 2003). This hypothesis was supported by evidence from the release of the puffer fish and zebrafish genome

sequences (Van de Peer *et al.,* 2003; Venkatesh, 2003). Genome duplication in the evolution of common carp is supported by the following observations. Its chromosome number (n = 100) is twice that of other Cyprinidae, and its DNA content is higher (Ohno, 1970). In addition, about 52% of this carp enzymes show a pattern consistent with duplication (Ferris and Whitt, 1977). Tetraploidization of carp was suggested to have taken place about 50 Million Years Ago (MYA), similarly to catastomids since both express a similar proportion of enzymes in duplicates. The c-myc genes in carp gave an estimate of 58 MYA for the event of tetraploidization (Zhang *et al.,* 1995). Other duplicated genes of the carp suggest a more recent divergence time of less than 16 MYA (Larhammar and Risinger, 1994).

The strongly supported grouping of SL sequences from cyprinids, catfish, rainbow trout and eel, combined with the robust association of carp SL with all other SL sequences (Fig. 2), supported the existing evidence of the origin of *sl*α and *sl*β after the last round of tetraploidization, with the subsequent maintenance of both paralogous SL copies in cyprinids and rainbow trout. It has been proposed that tetraploidizations generally lead to a higher survival of duplicates compared with small-scale gene duplications because thereby the balance in whole cell-biological pathways ("modules") may be maintained (Birchler and Veitia, 2010). Notably, *sl*β was subsequently lost from the most diverse and species-rich group of teleost fishes, the spiny-rayed fishes (Ocampo Daza and Larhammar, 2018). It is interesting that all fish that had the SL gene live at least a significant portion of their lives in fresh-water. Duplication of SL genes in those species may suggest important roles for SLs in an ancestral fish that moved into fresh water and required a second SL gene to help maintain homeostasis, including ion balance, body coloration and pigmentation, all functions associated to SL (Fukamachi *et al.,* 2004).

slα and *slβ* transcript sequences isolated in this work include 5`- and 3`-UTRs (Fig. 6), which means that for the first time the complete sequences were revealed because last published data on carp lacked UTR sequences (Valenzuela *et al.,* 2015) and were not even available from any other species where SL was sequenced (Ono *et al.,* 1990; Iraqi *et al.,* 1993; Pendón *et al.,* 1994; Astola *et al.,* 1996; Amemiya *et al.,* 1999). Therefore these data are of great relevance to deepen the understanding of the regulation of *sl* genes, because in particular UTRs are rich in regulatory elements and features shaping control of gene expression at the post-transcriptional level (Mignone *et al.,* 2002; Leppek *et al.,* 2018; Mayr, 2018;).

SLα and SLβ derived aminoacid sequences did not show high identity score (46,2%), but retained exact number and almost exact position of cysteine residues, just differing position of the third cysteine outside the predicted signal peptide (Fig. 7). Cysteine is critical for the formation of disulfide bond for all proteins as well as for specific binding of hormones to receptors (Sinha, 1995). Although SL belongs to the same superfamily and is structurally and functionally similar to GH and PRL, the conserved cysteines in SLs are different in number and position from those of GH and PRL. There are four in GH and PRL, which can form two disulfide bonds (Lynn and Shepherd, 2007), while there are eight in SLs of flounder and Atlantic cod, of which six could form three disulfide bonds (Ono *et al.,* 1990; Rand-Weaver *et al.,* 1991). Other SLs characterized in different fish have different cysteine number, such as eel (*Anguilla anguilla*), Atlantic salmon, and Mozambique tilapia (May *et al.,* 1997; Benedet *et al.,* 2008; Uchida *et al.,* 2009), in general there are more than six cysteines in SL pre-peptides. In the present work, eight cysteines were found in carp SLα and SLβ predicted proteins. The last six cysteines present in the mature peptide are conserved in position and number between SLα and

SLβ, except for the third in the mature peptide, that its slightly displaced, similar to the situation of SLs found in *D. rerio, C. idella* and *G. rarus* (Zhu *et al.,* 2004). In other words, this difference between cysteine positions in SLα and SLβ is conserved when comparing species (Liu *et al.,* 2018). As was previously described in atlantic cod, three disulfide bonds are formed among the first two and last four cysteines within the SL molecules (Rand-Weaver *et al.,* 1991). This suggests that, despite the poor identity score, the overall three dimensional structure of SLα and SLβ might be very similar, which should be the case for isoforms of this hormone. But since the second disulfide bond formed between third and fourth cysteine occurs in the mature SLβ peptide, this suggests that there should be some difference in structure and therefore in function between SLα and SLβ. This specific aspect of SLβ appears conserved across species which suggests a similar functional development of the new SLβ variant in evolutionary more recent species.

Glycosylation is a well-known post-translational modification for most functional proteins, and there are mainly two kinds, N-glycosylation and O-glycosylation (Suga *et al.,* 2018). One putative N-glycosylation site was found on carp SLα and SLβ derived aminoacidic sequences, two putative O-glycosylation sites on SLα and one on SLβ (Fig. 7). As was previously described in cod there are two SL variants, a smaller non-glycosylated (23 kDa) and a bigger glycosylated variant (26 kDa). Cod SL has two possible N-glycosylation sites, but only one appears to have carbohydrate units attached. Chemical analysis showed the following sugars to be present: galactose, mannose, N-acetylneuramic acid and glucosamine (Rand-Weaver *et al.,* 1991). Also in grass carp were described two SL variants of 27 kDa and 29 kDa (Jiang *et al.,* 2008). According to the multiple sequence alignment, the SLα N-glycosylation site at residue number 145 (NKT) is conserved in most

fish SLs, however, the corresponding site in SLβ is different (NRS). This differences between SLα and SLβ is conserved in other species such as rare minnows and grass carp (Liu *et al.,* 2018). With the hypothesis that glycosylated and non-glycosylated SLs may have different physiological roles (Kakizawa *et al.,* 1993), the divergence of SLβ in this site provided another evidence to the viewpoint that SLβ has acquired a different set of functions from those of SLα in fish (Yang and Chen, 2003; Zhu *et al.,* 2004). Immunodetection of carp SLα, with a specific antibody developed during this work, also showed two different immunoreactive bands, similar to the pattern found for grass carp, which might correspond to carp SLα variants, a glycosylated and a non-glycosylated protein (Fig. 8).

After analyses of *in silico* predicted characteristics and size variants, as next step to clarify differences between both *sl* genes, the spatial expression pattern was determined and transcript quantification of both SL was carried out (Fig. 9). SL immunodetection showed reaction in cell groups located in neurointermediate lobe, whereas clearly the antibody specific for anti-carp-SLα displayed signal in just a portion of the region of these cell groups. With this experiment was shown that there are regions with SLβ producing cells where at the same time no cells with SLα immunoreaction were detected. Although the histochemical detection is of limited resolution, clearly different immunoreactive regions were distinguished when applying anti-SL or anti-SLα antibodies. The distinct local expression might be related to a subfunctionalization in a particular cell region in the neurointermediate lobe, a region interwoven with projections of neuronal cells from hypothalamus forming the particular hypothalamus-hypophyseal connection in fish, which lacks the mammalian specific portal system (Montefusco-Siegmund *et al.,* 2006). In earlier studies *sl* transcripts were detected in the *pars intermedia* (PI),

but absent from the other parts of pituitary gland which in fish in addition to PI is composed of segregated cell populations of lactotrophs in RPD and the GH expressing region in PPD (López *et al.,* 2001; Figueroa *et al.,* 2005) Corroborating differential spatial expression, similarly in *Danio rerio* using *in situ* hybridization, *SLα* and SLβ expression was detected in different sub-regions, SLα in the posterior part of the *pars intermedia* bordering the neural tissue that penetrates the *pars intermedia,* whereas SLβ was expressed in the anterior part of the *pars intermedia* bordering the *pars distalis* (Zhu *et al.,* 2004).

RNA quantified at basal conditions in male carp pituitary revealed significant difference of transcript level from *sl*α and *sl*β with lower levels of *sl*β respect to *sl*α (Fig. 9, B), in concordance with what was found in goldfish (Jiang and Wong, 2014), again confirming that both genes are expressed differentially. Taken together, differences in the cysteines, potential N-glycosylation sites and the spatio and quantitative expression profiles suggest a subfunctionalization of carp SLα and SLβ.

Since SL is such a multifunctional hormone in fish, it was shown that many factors are capable to disrupt or regulate its expression or secretion. This includes environmental stresses such as deprivation of light (Zhu *et al.,* 1999) and exposure to heavy metals (Valenzuela *et al.,* 2015; Liu *et al.,* 2018), or endogenous factors as corticotropin-releasing factor (CRF), gonadotropin-releasing hormone (GnRH), dopamine and serotonin (Kakizawa *et al.,* 1997). However, differential response of both *sl* genes was reported in response to autocrine stimulus (Jiang and Wong, 2013) and exogenous estrogen treatment (Valenzuela *et al.,* 2015). The present work focused on deepening the knowledge related to this last topic.

Since this differential regulation of the expression of *sl*, could be influenced by genetic or epigenetic mechanism, in a first step cis-regulatory elements were characterized. Therefore, identification of the transcription start site was of primary importance, since core regulatory elements are located around it, and in eukaryotic genes might be far up from the initiation of translation, sometimes even in another exon.

Genomic sequences obtained were compared with mRNA 5´-UTR sequences to identify the transcription start site, exons and introns. Exon-Intron distribution, length and number in both slα and slβ were similar to the data reported on chum salmon and sea bream (Astola *et al.,* 2004). Exon length and number seems to be well conserved between carp slα and slβ, however, there is mayor differences in intron length. Promoter regions appear to be of particular interest in the study of duplicate genes and have been addressed in several investigations. However, there is just one publication of the study of *sl* promoter, in sea bream, with the limitation of focusing on the study of only one of the genes and even without being able to determine the transcription start site (Astola *et al.,* 2004). This is why determination of exact transcription start site in both *sl* genes provides new and relevant information about *sl* expression regulation (Fig. 10).

Promoter regions of paralogs have been reported able to diverge, allowing expansion of functionality as well as changes in expression levels (Louis, 2007). As a result of these changes, complementary or novel organism functions could arise from differential regulatory mutations (Hellsten *et al.,* 2007). This work focused on promoter region, 500 bp upstream transcription start site. In both genes contains a consensus sequence for the TATA box element, numerous potential transcription factor binding sites. Pit-1 putative

binding sites were found on this region (Fig. 12 and Fig. 13), four on *sl*α and three on *sl*β, as expected because this transcription factor that is of paramount importance for the expression of GH-PRL superfamily genes and in consequence for setup and maintenance of the hypothalamus-pituitary axis (Dasen and Rosenfeld, 2001). The transcriptional regulation of Pit-1 on GH-PRL superfamily genes expression has been studied in several mammalian species, extensively in human and in lower vertebrates such as teleosts mainly in the model zebrafish (Dasen and Rosenfeld, 2001). Principally development of pituitary follows a similar cascade of Pit-1 binding sites have been described on GH promoter from sea bream (Almuly *et al.,* 2005) and in experiments of cotransfection into HeLa cells, rat Pit-1 enhanced the promoter activity of *sl* gene, and a 0.5 kilobase upstream region from the transcriptional start site was sufficient for this enhancement (Ono *et al.,* 1994). Despite the relevance of Pit-1 for SL expression, just one site was conserved in the same position, site 1 (in slα) and site 6 (in slβ) (Fig.27 and Table 9). This site is conserved amongst other species and even in other genes of the same superfamily as Growth hormone suggesting that this particular Pit-1 site is relevant for the expression of *sl* genes and probably for GH-PRL superfamily (Almuly *et al.,* 2005). Other putative binding sites were found on both genes, as SP-1, Glucocorticoid receptor and GATA, however also not conserved in number and localization between both genes.

Even more interestingly, putative binding site to estrogen receptor and a xenobiotic response element, were found on *sl*β promoter but not in *sl*α promoter (Fig. 27 and Table 9). These results might imply differential expression in response to estrogen, and will be addressed in depth later. Putative binding sites to Pit-1 and ER were assessed with mobility shift assay with recombinant protein and male carp pituitary nuclear extract. Confirming

three binding sites on each *sl* gene promoter (Fig.16). In these experiments, *in vitro* under controlled conditions intensity of interaction of nuclear extract containing Pit-1 with short DNA fragment permits to estimate affinity to slightly different sequences. Each of these particular sequences evolved through millions of years in the complex interplay of factors fine tuning adaptation to changing environment, therefore the analyses of different Pit-1 binding site constitutes a powerful evaluation how mutations in regulatory elements arise to gain functional relevance for gene expression regulation. Therefore, analyses of regulation of duplicated genes that evolved in the same cell in an organism constitute profound insights in natural forces shaping genome function.

With another approach site directed mutagenesis revealed that the first three proximal Pit-1 sites in trout are necessary for optimal gene activation. Moreover, Pit-1 binding to the second and third binding sites was shown to mediate the cAMP response in trout (Argenton *et al.,* 2002). Those sites displayed different binding affinities (Fig. 17), of which site 3 on slα promoter showed the higher affinity. Higher Pit-1 binding affinity was reported that is related with presence of CATT sequence (Agarwal and Cho, 2018) that its present on site 3 but no in the other sites. The fact that the *slα* promoter has the site with the highest affinity for Pit-1 correlated with higher basal *slα* expression levels. This is consistent with the pivotal role played by Pit-1 in the activation of its target genes and the idea that Pit-1, as other POU proteins, also plays a permissive role enabling the involvement of associated factors (Ryan and Rosenfeld, 1997). However, it does not explain differential regulation of *sl* genes in response to estrogen in male carp pituitary, taking into consideration that Pit-1 expression levels were not affected significantly

Name	Transcription factor	Gene	Position (respect to +1)	Sequence	Evidences
Site 1	Pit-1	slα	-41 to -50	AATATTCAC	*In silico* EMSA (extract and recombinant)
Site 2	Pit-1	slα	-62 to -71 -65 to -74	TAGATTCAT TTTATGAAT	*In silico* EMSA (extract and recombinant)
Site 3	Pit-1	slα	-187 to - 196	ATTATTGAT	*In silico* EMSA (extract and recombinant)
Site 5	Pit-1	slβ	-20 to -11	GTTATTCAT	*In silico* EMSA (extract and recombinant)
Site 6	Pit-1	slβ	-40 to -49 -42 to -52	TTTATGAAT TTAATTCAT	*In silico* EMSA (extract and recombinant)
Site 7	Pit-1	slβ	-122 to -131 -125 to -134	TTTATGCAT ATGATGCAT	*In silico* EMSA (extract and recombinant)
ER1	Estrogen Receptor	slβ	-100 to -105	GGTCAnnn	*In silico* EMSA (extract) Reporter assay
XRE	aryl hydrocarbon receptor	slβ	-228 to -234	CACGCT	*In silico*

Table 9. Transcription factor binding sites to both sl genes promoters.

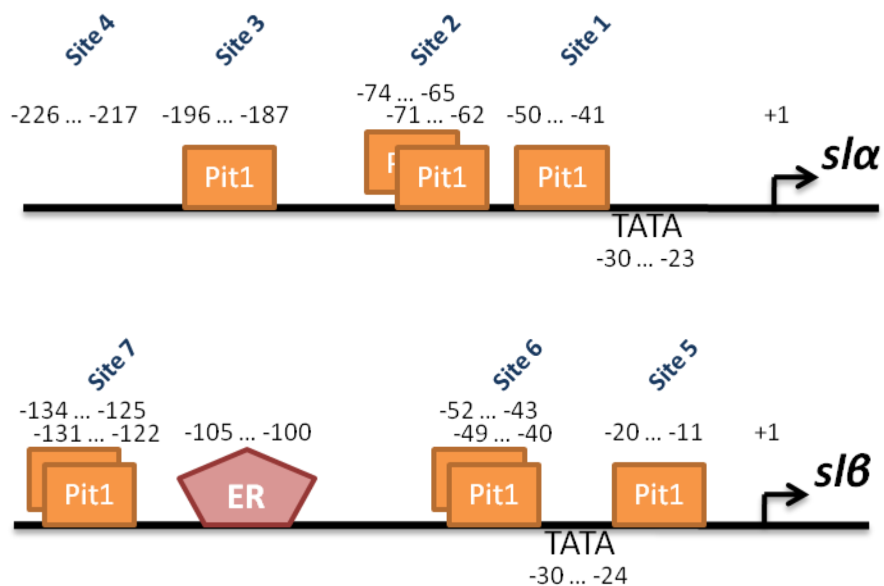


Figure 27: Carp *slα* and *slβ* proximal promoter scheme. Key landmarks as TATA box and transcription start site (+1) are represented here. Transcription factors binding sites confirmed by EMSAs, are depicted in orange (Pit-1) and pink (ER), numeration its related to +1 site.

by estrogen (Valenzuela *et al.,* 2015). So that a factor(s) other than the Pit-l may be important for *sl* differential response.

In general, both promoters shared relative low identity, and even CpG dinucleotide number and location are different (Fig. 14). In the context of the tetraploid carp fish, the presence of highly divergent regions in the regulatory sequences of the two copies of the *sl* genes embedded into otherwise well-conserved regions, is of particular interest (Ravi and Venkatesh, 2018). Well-illustrated examples exist for gene loss, pseudogenization and sub- or neo-functionalization (Postlethwait *et al.,* 2004). Sequence divergence of the two gene copies might affect the protein coding region, leading to a protein presenting slightly different biological properties, or the regulatory regions of the gene, leading to a modification of its spatial or temporal expression as in this case. The divergence of duplicate gene promoters in response to a specific stimuli has not been studied in depth, however has been reported in carp important differences on cis-regulatory elements from two functional Pit-1 genes (Kausel *et al.,* 2006), nevertheless in that case, Pit-1 genes promoter showed highly conserved regions immediately downstream TATA box. This is an interesting comparison, because two different duplicated genes showed different levels of divergences suggesting that duplicated genes fate, does not only depend on time after duplication or environmental input, but gene function or even position, and several variables more. Another important example in carp was described with the POMC gene, another gene differentially expressed in the pituitary (Arends *et al.,* 1998). All these data support that carp fish provides an excellent model system to study the dynamics of duplicated genes after genome duplication (Cossins and Crawford, 2005; Wendel *et al.,* 2018).

ER binding site located on *sl*β gene was confirmed by electromobility shift assay (Fig. 16 B). The biological effects of estrogens were mediated by binding to estrogen receptors. Through a variety of mechanisms and interactions with other transcription factors and distinct coactivators and corepressors, these binding events can result in either transactivation or transrepression of target gene expression. High levels those receptors are expressed in the pituitary and the hypothalamus (Stefaneanu *et al.,* 1994). After ligand binding, estrogen receptor form homodimers or heterodimers and classically bind to specific estrogen response elements in the promoter region of target genes (McKenna and O'Malley, 2000). The consensus estrogen response element (ERE) contains a palindromic sequence of two 5 bp separated by three random nucleotides, 5'-GGTCAnnnTGACC-3' (Klinge, 2001). In this case, interestingly ERE site found on *sl*β β corresponds to only half of this sequence, but in electromobility shift assay with pituitary extract showed binding reaction that disappeared when incubated with unlabelled probes containing canonical ERE, ERE half site, and a ERE site previously described for carp isotocin promoter. This suggests that ER can bind this site despite that it is just an ERE half site. Literature describe that ER can bind modified ERE including ERE half site, however sequence variations affect binding affinity (El Marzouk *et al.,* 2008). It was shown that EREs in which nucleotides were altered in each arm of the palindrome showed lower transcriptional activity than those containing alterations in only one half of the ERE palindrome (Klinge, 2001). The fact that *sl*β β ERE half site contains exactly the same sequence than consensus ERE half site, take relevance if we take into account that it was confirmed that ER has higher affinity for ERE half-sites than for the imperfect ERE (Anderson and Gorski, 2000). It has also been reported in heterologous Rat cell lines, both Pit-1 and ER must be

present for estrogen regulation of the prolactin gene to occur (Kaneko *et al.,* 1993), and that the proximity of a Pit-1 binding site to an perfect ERE half site improved their binding affinity (Murdoch *et al.,* 1995). Therefore, our data confirm the importance of location of Pit-1 site respect to ERE for sensitivity to estrogen when regulating expression of the target gene. Indeed, in the case of *sl*β β promoter, a Pit-1 site (Pit-1 site 7) is located close to ERE half site and was confirmed to bind with EMSA (Fig. 13).

After showing the binding of nuclear extract factors *in vitro* and in order to ensure that this particular ERE in *sl*β promoter is related to differential regulation of *sl*β in response to estrogen, an *in vivo* strategy was employed, using Luciferase reported assay. Transient transfection into B6/GH3 rat pituitary cells with a plasmid containing firefly luciferase under control of wild type *sl*β promoter (pGL3/SLB) and *sl*β promoter with mutated ERE binding site
(pGL3/SLBEREmut) exposed to estrogen or vehicle, showed that cells transfected with pGL3/SLB construct increase luciferase activity in response to estrogen, however, cells transfected with pGL3/SLBEREmut construct lost the ability to respond to estrogen. This suggests that this particular ERE its directly related with the differential response to estrogen of *sl*β gene in male carp pituitary previously reported (Valenzuela *et al.,* 2015). There is strong evidence of direct relationship of ERE and regulation of expression of GH and PRL, however, it focuses on studies in mammals (Fujimoto *et al.,* 2004; Avtanski *et al.,* 2014). In this study, for the first time, a direct relationship is established between a specific ERE and the differential response of duplicated genes to an external stimulus such as estrogen in a teleost fish (Fig. 28).

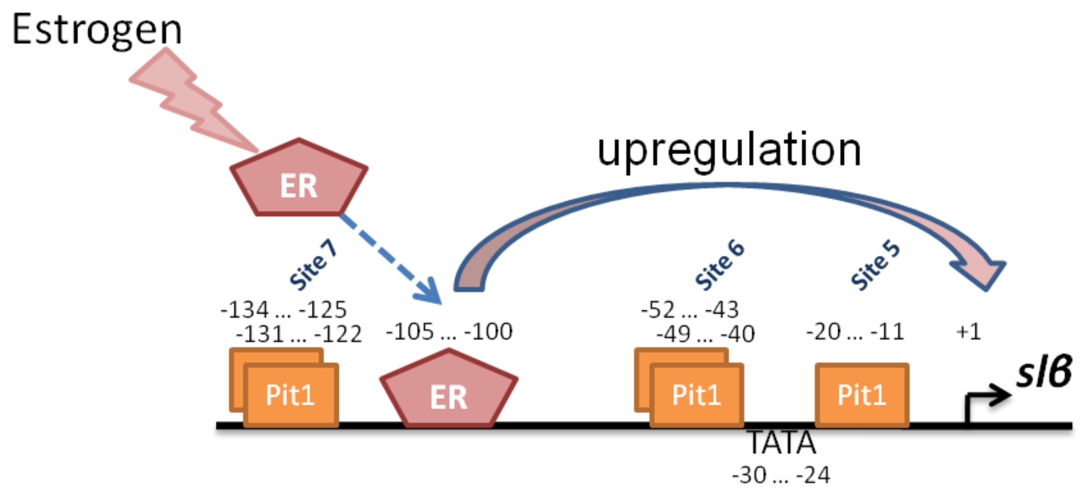


Figure 28: Suggested *slβ* estrogenic response regulation. Key landmarks as TATA box and transcription start site (+1) are represented here. Transcription factors binding sites confirmed by EMSAs, are depicted in orange (Pit-1) and pink (ER), numeration its related to +1 site.

In addition for the first time a putative Xenobiotic Response Element (XRE), was also found on *sl*β promoter, but no in *sl*α (Fig. 13). XRE, is a core nucleotide sequence upstream of inducible target genes for the transcription factor aryl hydrocarbon receptor (AhR) that is responsible for signal transduction of exogenous environmental pollutants in eukaryotic cells (Kuramoto *et al.,* 2003; Gore *et al.,* 2015). This could explain differential response of *sl*β in pituitary of male carp treated with Benzo-a-Pyrene (BaP), a member of Polycyclic Aromatic Hydrocarbons compounds (Fig. S. 1), opening the possibilities to new research on the divergence of the regulatory regions of these duplicated genes in response to external stimuli.

Anthropogenic activity can influence teleost physiology indirectly, through the modification of naturogenic environmental factors, or directly in the form of aquatic contaminants, such as estrogens. These environmental factors can elicit epigenetic changes in teleost fish, which contribute to shaping genetic and in consequence physiological responses across different timescales. To investigate the possible relationship between differential response of *sl* genes to estrogen and epigenetic marks, methylation pattern and chromatin compaction state was assessed (Fig. 20).

DNA methylation changes in response to anthropogenic contaminants have been widely studied in teleost fish, and have largely focused on determining context-dependent changes in expression profiles of dnmt paralogues, Dnmt activity and differential global or specific DNA methylation, the latter of which have been assessed both genome-wide and in promoter regions of specific genes. These studies have been conducted mainly in the zebrafish model. Novel approaches involving next generation sequencing were used previously to identify CpG-rich regions in various studies in fish (Masser *et al.,*

2015; Koganti *et al.,* 2017). Although estrogen did not affect CpG methylation substantially, various reports have confirmed site-specific DNA methylation as a mechanism in regulating gene expression (Ziller *et al.,* 2013). CpG methylation in promoter regions largely downregulates gene expression while most CpG methylations in the coding region enhance gene expression, although it was also demonstrated that intragenic CpG frequency and distribution impacts transgene and genomic gene expression levels in mammalian cells (Bauer *et al.,* 2010; Krinner *et al.,* 2014). The presence of CpG and non-CpG sites neighboring ERE suggests probable interaction between these two mechanisms to regulate *sl* gene expression (Fig. 13), however no significant changes in methylation of neither CpG or CpH sites were detected, at least in this time point, in response to estrogen in both *sl* promoter and coding regions (Fig 23 and 24). As was previously reported in rainbow trout, CpG sites on regulatory regions of MyoD gene are not differentially methylated in response to estrogen treatment, what is consistent with the results of this work. However, elevated CpH methylation in estrogen treated samples suggested an effect of estrogen on *de novo* methylation of MyoD *in vivo* (Koganti *et al.,* 2017), which is unlike what was found in carp *sl* genes. Interestingly the general levels of CpG methylation on *sl*α promoter are drastically lower than *sl*β levels (Fig. 23 A and 24 A) what is consistent with basal expression levels of both genes (Fig. 9). This suggesting that CpG methylation is somehow related in maintenance of basal differential expression levels of both *sl* genes, which is very interesting since it has been reported that comparative genome analysis regarding duplicate genes supports the hypothesis that differential DNA methylation and epigenetic changes play a role in protecting duplicate genes from pseudogenization (Rodin *et al.,* 2005; Cortese *et al.,* 2008). However, there is no evidence about a relationship with

estrogen response, at least in assessed conditions, inquiries in other temporary windows remain to be done. Collectively, these observations improve our understanding of the effects of estrogen on DNA methylation in the context of duplicated gene in teleost.

Transcriptional activity and structure of chromatin are correlated with patterns of covalent DNA and histone modification. In order to gain insight of how chromatin state around both *sl* genes is affected by estrogen treatment ChIP experiments were carried out, showing a decrease of heterochromatin mark on *sl*β promoter region in response to estrogen in male carp pituitary and no changes in euchromatin or heterochromatin mark in *sl*α promoter. These results are consistent with previously reported differential regulation of both *sl* genes in response to estrogen. Similar results were reported for L41 gene when the transcriptional activity of this gene is seasonally repressed (Simonet *et al.,* 2013). Teleost chromatin goes through important modulation during adaptation to an exogenous stimulus as histone posttranslational modification (Bannister and Kouzarides, 2011) and histone variants (Araya *et al.,* 2010), suggesting that the epigenetic regulation in this species constitutes a complex mechanism. It is important to note that the results of the epigenetic analyzes of both genes of SL are not conclusive, this may be due to the fact that they were made in extracts of complete pituitary, which is composed of a variety of cell types, many of which do not express SL, with the consequent epigenetic effect related to the silencing of genes, which could make a buffering effect of the observed results. However, in turn, they give more importance to the observed differences, which must be sufficiently robust to be appreciated despite the aforementioned technical drawbacks.

For the first time regulatory elements were revealed that had evolved to control similar sequences but in different genome locations from two *sl* genes in carp which support the idea that intricate genetic and epigenetic divergence of duplicate genes affects gene expression and functional divergence of duplicate genes. The teleost genome duplication presents important advantages for the analysis of gene function and fate, because of principles that govern the evolution of gene duplicates. After genome duplication, each gene copy can follow a separate evolutionary trajectory, in the case of carp *sl*, take it to a process of subfunctionalization at the level of cis regulatory and epigenetic elements, which impact on their location and basal expression levels and in response to external stimuli.

Undoubtedly, the novel insights about *sl*β's versatile response to estrogen should be taken into account, when developing biomarkers for monitoring of environmental endocrine disrupting effects, where evaluation of *sl*β expression could indicate early alert against potentially harmful changes in the aquatic environment.

7. CONCLUSIONS

- Two **somatolactin** transcripts, ***slα* and *slβ***, were detected in carp pituitary in basal condition with higher mRNA levels of *slα* respect to *slβ*; at the protein level SL was immunodetected with anti-SL antibody in *pars intermedia* (PI) and with specific anti-SLα immunoreaction occurred in a subregion of PI of pituitary gland; the **quantitative and spatially different expression pattern** suggests differential regulation of expression of *slα* and *slβ* genes.
- The derived amino acid sequences of SLα and SLβ share less than 50% identity, but contain key features such as cysteines conserved in number and position, nevertheless the third cysteine in the mature peptide that its slightly displaced, this difference is conserved when comparing species and suggests some difference in structure and therefore in function between SLα and SLβ.
- Between *in silico* cloned *slα* and *slβ* genes, number and length of exons is conserved, also in comparison with other species; however, length of introns differ substantially; more important, *slα* and *slβ* gene sequences were found in **different contigs consistent** with a **common ancestral sequence** before the origin of the **duplicate gene sequences** evolving in the **polyploid carp genome**.
- For *slα* as well as for *slβ* gene a **unique transcription start site** was determined unequivocally by 5'-RACE, a TATA-box at position -30 and other conserved cis-regulatory elements.
- In the overall highly divergent promoter regions were predicted several consensus binding sites for **Pit-1**, the **master regulator** for development and function of pituitary SL producing cells. In both *sl*

genes the best conserved **Pit-1 binding site** is located in the neighborhood of TATA box which might indicate a prominent regulatory role for transcription initiation.

- From the predicted Pit-1 binding sites, in each of the *sl*α and *sl*β promoter region three sites were confirmed with *in vitro* binding assays, highest affinity was determined for a Pit-1 site in *sl*β. In addition, specific interaction of nuclear extract *in vitro* suggested a functional role for the **ERE-site only** present **in *sl*β** but not in *sl*α promoter.
- Indeed, only *sl*β transcripts increased significantly in pituitary of adult carp in response to estrogen treatment and the reporter assay in pituitary derived cell line clearly revealed that the ERE-site is necessary and sufficient for transcriptional response to estrogen *in vivo*. Therefore, *sl*β seems to have acquired **novel regulatory gene elements for subfunctionalization in a spatio – estrogen stimulus dependent manner**.
- When profiling epigenetic characteristics, no differences were found when comparing CpG methylation pattern of promoter region to coding region, this was the case in *sl*α as well as in *sl*β gene. However, *sl*α promoter methylation was lower than in *sl*β promoter in basal condition, correlating with the higher transcript level of *sl*α with respect to *sl*β in adult carp pituitary suggesting a role for DNA methylation in regulation of *sl* expression.
- According to the estrogen elicited transcriptional response of *sl*β but not of *sl*α, only in the promoter of *sl*β showed a decrease on of inactive chromatin marks; these data suggest **divergence of epigenetic regulatory pathways modulating expression of these duplicated genes**.

Taken together these results showed that promoter regions of *sl* paralogs diverged, promoting an expansion in regulatory capabilities, in this case conferring one of the genes the new ability to respond to estrogen. Therefore, herewith is clearly shown that the carp with its duplicated genome constitutes an excellent tool to study the divergence of regulation of expression of duplicated genes both at genetic and epigenetic level during millions of years in the complex system of an adult organism.

This study has identified relevant genetic and epigenetic events important for differential regulation of duplicated SL genes and provides novel insights into understanding regulatory mechanisms of fish to estrogen, a widespread endocrine disrupting compound in the aquatic environment.

8. SUPPLEMENTARY FIGURES

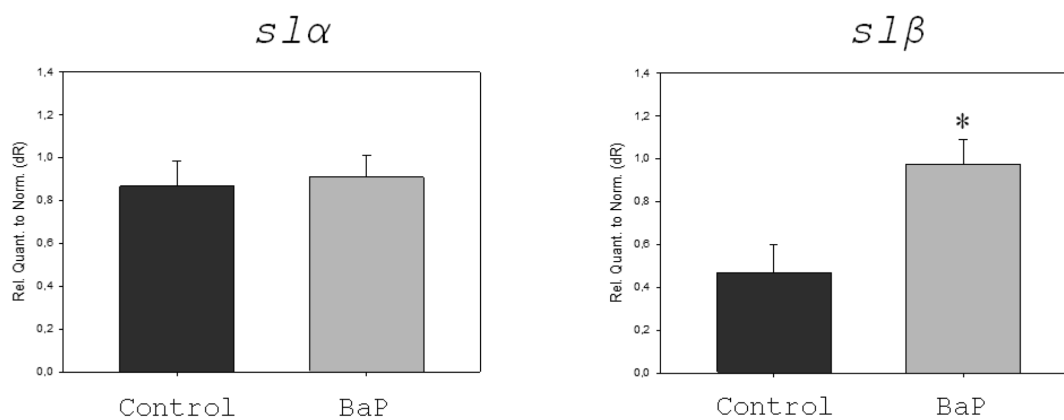
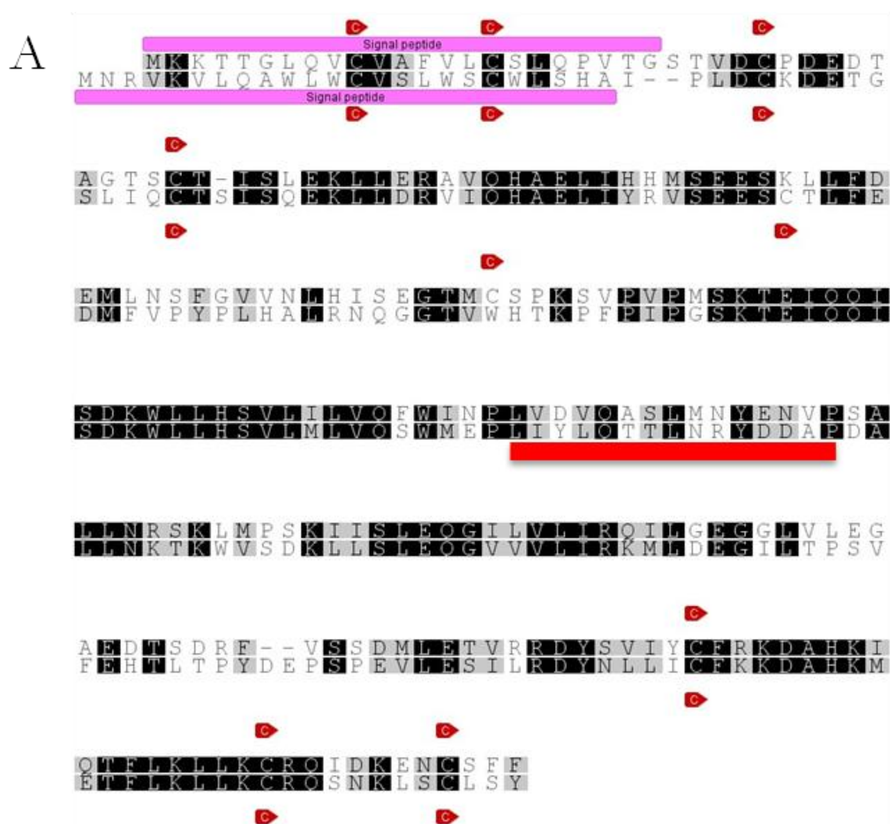


Figure S 1. Effect of BaP on pituitary *sl* gene expression. The graphs show relative quantification to the normalizer gene (dR) of RT-qPCR analysis of (A) slα and (B) slβ expression in total pituitary RNA (control n = 4; BaP n = 4). The bars indicate the standard deviation, (*) Student's T test, $P < 0.05$ was considered significant.



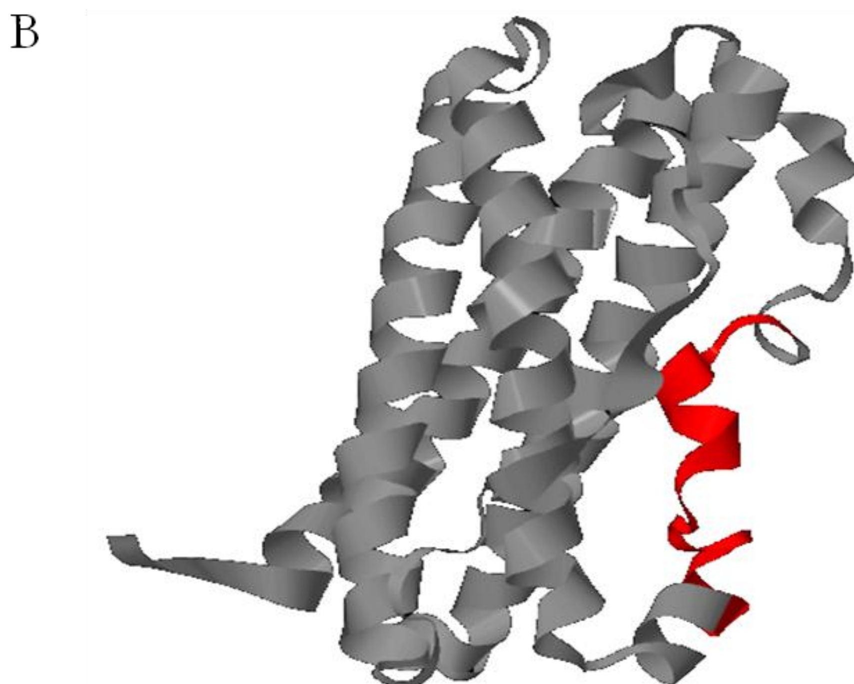


Figure S 2 Epitope selection for carp slα antibody. (A)Amino acid sequence alignment of two SL carp genes. SLα 234 aa (Acc. Number:

ADE60529.2) and SLβ 230 aa (Acc. Num: AAY45791.2) Similarity is represented in grey scale. Darkest residues represent strong similarity. Cysteine residues are remarked with red arrows. Signal peptide are highlighted in pink and Peptide used for antibody production is underlined in red.(B) Predicted SLα structure, Homology 3D model produced using SWISS-MODEL tool, based on SLα derived aminoacidic sequence (Acc. Number: ADE60529.2). Peptide used for antibody production is depicted in red.

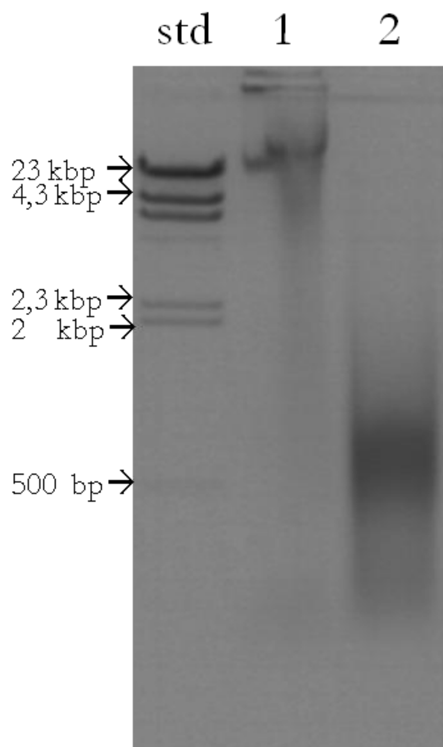


Figure S 3. gDNA fragmentation for immunoprecipitation. Image shows 1,2% agarose gel electrophoresis of gDNA from male carp pituitary obtained after chromatin extraction and subsequent mechanical fragmentation by sonication. std correspond to Lambda/*Hin*dIII marker (Thermo Fisher Scientific), lane 1 gDNA before sonication, lane 2 gDNA after sonication.

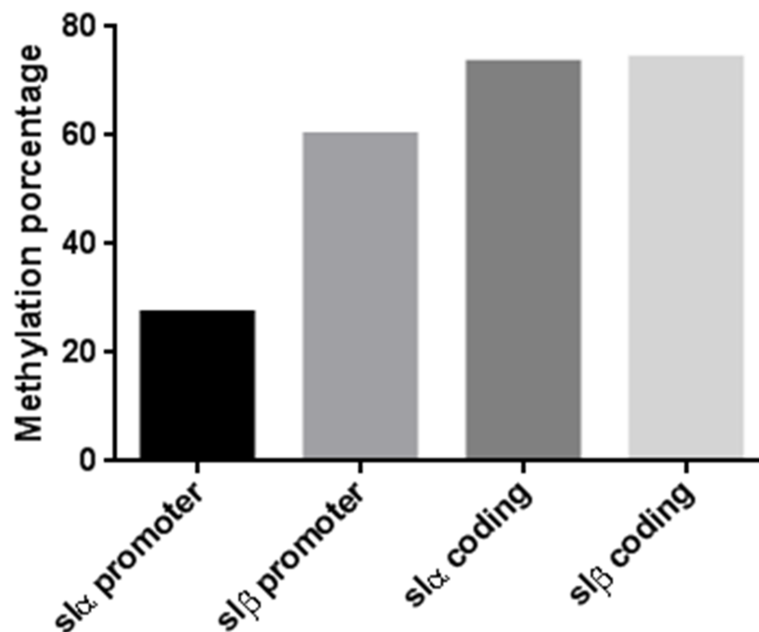


Figure S 4. Methylation levels on both *sl* genes under basal conditions. Graph depict general methylation percentage of both *sl* genes, taking into account all the sites analyzed in each region (Pooled gDNA from 4 individuals, 12 clones sequenced for each region) in adult male carp pituitary under basal conditions without exogenous treatment.

9. BIBLIOGRAPHY

Agarwal, S., and Cho, T. Y. (2018). Biochemical and structural characterization of a novel cooperative binding mode by Pit-1 with CATT repeats in the macrophage migration inhibitory factor promoter. *Nucleic Acids Res.* 46, 929–941. doi:10.1093/nar/gkx1183.

Allis, C. D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* 17, 487–500. doi:10.1038/nrg.2016.59.

Almuly, R., Poleg-Danin, Y., Gorshkov, S., and Gorshkova, G. (2005). Characterization of the 5' flanking region of the growth hormone gene of the marine teleost, gilthead sea bream Sparus aurata: analysis of a polymorphic …. *Fish. Sci.* 71, 479–490.

Ambrosi, C., Manzo, M., and Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. *J. Mol. Biol.* 429, 1459–1475. doi:10.1016/j.jmb.2017.02.008.

Amemiya, Y., Sogabe, Y., Nozaki, M., Takahashi, A., and Kawauchi, H. (1999). Somatolactin in the white sturgeon and African lungfish and its evolutionary significance. *Gen. Comp. Endocrinol.* 114, 181–190. doi:10.1006/gcen.1998.7250.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711–1714. doi:10.1126/science.282.5394.1711.

Anderson, I., and Gorski, J. (2000). Estrogen receptor alpha interaction with estrogen response element half-sites from the rat prolactin gene. *Biochemistry* 39, 3842–3847.

Araya, I., Nardocci, G., Morales, J., Vera, M., Molina, A., and Alvarez, M. (2010). MacroH2A subtypes contribute antagonistically to the transcriptional regulation of the ribosomal cistron during seasonal acclimatization of the carp fish. *Epigenetics Chromatin* 3, 14. doi:10.1186/1756-8935-3-14.

Arends, R. J., van der Gaag, R., Martens, G. J., Wendelaar Bonga, S. E., and Flik, G. (1998). Differential expression of two pro-opiomelanocortin mRNAs during temperature stress in common carp (Cyprinus carpio L.). *J. Endocrinol.* 159, 85–91.

Argenton, F., Vianello, S., Bernardini, S., Lopreiato, R., Colombo, L., and Bortolussi, M. (2002). Trout GH promoter analysis reveals a modular pattern of regulation consistent with the diversification of GH gene control and function in vertebrates. *Mol. Cell. Endocrinol.* 189, 11–23. doi:10.1016/S0303-7207(01)00757-2.

Arsovski, A. A., Pradinuk, J., Guo, X. Q., Wang, S., and Adams, K. L. (2015). Evolution of Cis-Regulatory Elements and Regulatory Networks in Duplicated Genes of Arabidopsis. *Plant Physiol.* 169, 2982–2991. doi:10.1104/pp.15.00717.

Arukwe, A., and Goksøyr, A. (1997). Changes in three hepatic cytochrome P450 subfamilies during a reproductive cycle in turbot (Scophthalmus maximus L.). *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology.*

Astola, A., Calduch-Giner, J. A., Ortiz, M., Pérez-Sánchez, J., and Valdivia, M. M. (2004). Genomic structure and functional analysis of promoter region of somatolactin gene of sea bream (Sparus aurata). *Mar. Biotechnol.* 6, 411–418. doi:10.1007/s10126-004-3210-2.

Astola, A., Pendón, C., Ortíz, M., and Valdivia, M. M. (1996). Cloning and Expression of Somatolactin, a Pituitary Hormone Related to Growth Hormone and Prolactin from Gilthead Seabream,Sparus aurata. *Gen. Comp. Endocrinol.* 104, 330–336. doi:10.1006/gcen.1996.0178.

Avtanski, D., Novaira, H. J., Wu, S., Romero, C. J., Kineman, R., Luque, R. M., Wondisford, F., and Radovick, S. (2014). Both estrogen receptor α and β stimulate pituitary GH gene expression. *Mol. Endocrinol.* 28, 40–52. doi:10.1210/me.2013-1245.

Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395. doi:10.1038/cr.2011.22.

Benedet, S., Björnsson, B. T., Taranger, G. L., and Andersson, E. (2008). Cloning of somatolactin alpha, beta forms and the somatolactin receptor in Atlantic salmon: seasonal expression profile in pituitary and ovary of maturing female broodstock. *Reprod Biol Endocrinol* 6, 42. doi:10.1186/1477-7827-6-42.

Bircher, S. (2011). *Phytoremediation of Natural and Synthetic Steroid Growth Promoters Used in Livestock Production by Riparian Buffer Zone Plants*. University of Iowa.

Birchler, J. A., and Veitia, R. A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 186, 54–62. doi:10.1111/j.1469-8137.2009.03087.x.

Blanc, G., and Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* 16, 1679–1691. doi:10.1105/tpc.021410.

Bolong, N., Ismail, A. F., Salim, M. R., and Matsuura, T. (2009). A review of the effects of emerging contaminants in wastewater and options for their removal. *Desalination* 239, 229–246. doi:10.1016/j.desal.2008.03.020.

Bromage, N., and Cumaranatunga, R. (1988). "Egg production in the rainbow trout," in *Recent advances in aquaculture*, eds. J. F. Muir and R. J. Roberts (Dordrecht: Springer Netherlands), 63–138. doi:10.1007/978-94-011-9743-4_2.

Caccavelli, L., Manfroid, I., Martial, J. A., and Muller, M. (1998). Transcription factor AP1 is involved in basal and okadaic acid-stimulated activity of the human PRL promoter. *Mol. Endocrinol.* 12, 1215–1227. doi:10.1210/mend.12.8.0152.

Carnevali, O., Santangeli, S., Forner-Piquer, I., Basili, D., Maradonna, F. (2018). Endocrine-disrupting chemicals in aquatic environment: what are the risks for fish gametes? *Fish Physiol. Biochem.* doi: 10.1007/s10695-018-0507-z.

Coe, T. S., Söffker, M. K., Filby, A. L., Hodgson, D., and Tyler, C. R. (2010). Impacts of early life exposure to estrogen on subsequent breeding behavior and reproductive success in zebrafish. *Environ. Sci. Technol.* 44, 6481–6487. doi:10.1021/es101185b.

Colborn, T., vom Saal, F. S., and Soto, A. M. (1993). Developmental effects of endocrine-disrupting chemicals in wildlife and humans. *Environ. Health Perspect.* 101, 378–384.

Conant, G. C., and Wolfe, K. H. (2006). Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biol.* 4, e109. doi:10.1371/journal.pbio.0040109.

Cortese, R., Krispin, M., Weiss, G., Berlin, K., and Eckhardt, F. (2008). DNA methylation profiling of pseudogene-parental gene pairs and two gene families. *Genomics* 91, 492–502. doi:10.1016/j.ygeno.2008.02.004.

Cossins, A. R., and Crawford, D. L. (2005). Fish as models for environmental genomics. *Nat. Rev. Genet.* 6, 324–333. doi:10.1038/nrg1590.

Cunliffe, V. T. (2016). Histone modifications in zebrafish development. *Methods Cell Biol.* 135, 361–385. doi:10.1016/bs.mcb.2016.05.005.

Dasen, J. S., and Rosenfeld, M. G. (2001). Signaling and transcriptional mechanisms in pituitary development. *Annu. Rev. Neurosci.* 24, 327–355. doi:10.1146/annurev.neuro.24.1.327.

Diamanti-Kandarakis, E., Bourguignon, J.-P, Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., … Gore, A.C. (2009). Endocrine-disrupting Chemicals: an Endocrine Society Scientific Statement. *Endocrine Reviews* 30, 293-342.

El Marzouk, S., Gahattamaneni, R., Joshi, S. R., and Scovell, W. M. (2008). The plasticity of estrogen receptor-DNA complexes: binding affinity and specificity of estrogen receptors to estrogen response element half-sites separated by variant spacers. *J. Steroid Biochem. Mol. Biol.* 110, 186–195. doi:10.1016/j.jsbmb.2008.03.034.

Fenske, M., Maack, G., Schäfers, C., and Segner, H. (2005). An environmentally relevant concentration of estrogen induces arrest of male gonad development in zebrafish, Danio rerio. *Environ. Toxicol. Chem.* 24, 1088–1098.

Ferris, S. D., and Whitt, G. S. (1977). Duplicate gene expression in diploid and tetraploid loaches (Cypriniformes, Cobitidae). *Biochem Genet* 15, 1097–1112.

Figueroa, J., Martín, R. S., Flores, C., Grothusen, H., and Kausel, G. (2005). Seasonal modulation of growth hormone mRNA and protein levels in carp pituitary: evidence for two expressed genes. *J Comp Physiol B, Biochem Syst Environ Physiol* 175, 185–192. doi:10.1007/s00360-005-0474-4.

Figueroa, J., Molina, A., Alvarez, M., Villanueva, J., Reyes, A., León, G., and Krauskopf, M. (1994). Prolactin gene expression and changes of prolactin pituitary level during the seasonal acclimatization of the carp. *Comp Biochem Physiol Biochem Mol Biol* 108, 551–560.

Filby, A. L., Paull, G. C., Searle, F., Ortiz-Zarragoitia, M., and Tyler, C. R. (2012). Environmental estrogen-induced alterations of male aggression and dominance hierarchies in fish: a mechanistic analysis. *Environ. Sci. Technol.* 46, 3472–3479. doi:10.1021/es204023d.

Fisher, R. A. (1935). The sheltering of lethals. *Am. Nat.* 69, 446–455. doi:10.1086/280618.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

Fujimoto, N., Igarashi, K., Kanno, J., and Inoue, T. (2004). Identification of estrogen-responsive genes in the GH3 cell line by cDNA microarray analysis. *J. Steroid Biochem. Mol. Biol.* 91, 121–129. doi:10.1016/j.jsbmb.2004.02.006.

Fukamachi, S., and Meyer, A. (2007). Evolution of receptors for growth hormone and somatolactin in fish and land vertebrates: lessons from

the lungfish and sturgeon orthologues. *J. Mol. Evol.* 65, 359–372. doi:10.1007/s00239-007-9035-7.

Fukamachi, S., Sugimoto, M., Mitani, H., and Shima, A. (2004). Somatolactin selectively regulates proliferation and morphogenesis of neural-crest derived pigment cells in medaka. *Proc. Natl. Acad. Sci. USA* 101, 10661–10666. doi:10.1073/pnas.0401278101.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318.

Goll, M. G., and Halpern, M. E. (2011). DNA methylation in zebrafish. *Prog Mol Biol Transl Sci* 101, 193–218. doi:10.1016/B978-0-12-387685-0.00005-6.

Gore, A.C., Chappell, V.A., Fenton, S.E., Flaws, J.A., Nadal, A., Prins, G.S., Toppari, J., Zoeller, R.T. (2015). Executive Summary to EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocrine Reviews* 36, 593-602.

Gu, Y. Q., Anderson, O. D., Londeorë, C. F., Kong, X., Chibbar, R. N., and Lazo, G. R. (2003). Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. *Genome* 46, 1084–1097. doi:10.1139/g03-071.

Hallgren, P., Nicolle, A., Hansson, L.-A., Brönmark, C., Nikoleris, L., Hyder, M., and Persson, A. (2014). Synthetic estrogen directly affects fish biomass and may indirectly disrupt aquatic food webs. *Environ. Toxicol. Chem.* 33, 930–936. doi:10.1002/etc.2528.

He, X., and Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169, 1157–1164. doi:10.1534/genetics.104.037051.

Hellsten, U., Khokha, M. K., Grammer, T. C., Harland, R. M., Richardson, P., and Rokhsar, D. S. (2007). Accelerated gene evolution and

subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Biol.* 5, 31. doi:10.1186/1741-7007-5-31.

Hewitt, L. M., Kovacs, T. G., Dubé, M. G., MacLatchy, D. L., Martel, P. H., McMaster, M. E., Paice, M. G., Parrott, J. L., van den Heuvel, M. R., and van der Kraak, G. J. (2008). Altered reproduction in fish exposed to pulp and paper mill effluents: roles of individual compounds and mill operating conditions. *Environ. Toxicol. Chem.* 27, 682–697. doi:10.1897/07-195.

Hittinger, C. T., and Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449, 677–681. doi:10.1038/nature06151.

Iraqi, F., Gong, Z., Hew, C. L., and Crim, L. (1993). Isolation and characterization of somatolactin genes from two cold water marine teleosts, lumpfish (Cyclopterus lumpus) and halibut (Hippoglossus hippoglossus). *Mol Marine Biol Biotechnol* 2, 96–103.

Jacq, C., Miller, J. R., and Brownlee, G. G. (1977). A pseudogene structure in 5S DNA of Xenopus laevis. *Cell* 12, 109–120.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33 Suppl, 245–254. doi:10.1038/ng1089.

Jiang, Q., Ko, W. K. W., Lerner, E. A., Chan, K. M., and Wong, A. O. L. (2008). Grass carp somatolactin: I. Evidence for PACAP induction of somatolactin-alpha and -beta gene expression via activation of pituitary PAC-I receptors. *Am. J. Physiol. Endocrinol. Metab.* 295, E463–76. doi:10.1152/ajpendo.90385.2008.

Jiang, Q., and Wong, A. O. L. (2013). Signal transduction mechanisms for autocrine/paracrine regulation of somatolactin-α secretion and synthesis in carp pituitary cells by somatolactin-α and -β. *Am. J. Physiol. Endocrinol. Metab.* 304, E176–86. doi:10.1152/ajpendo.00455.2012.

Jiang, Q., and Wong, A. O. L. (2014). Somatostatin-28 inhibitory action on somatolactin-α and -β gene expression in goldfish. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 307, R755–68. doi:10.1152/ajpregu.00193.2014.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi:10.1038/nrg3230.

Joseph, S. R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., and Vastenhouw, N. L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife* 6. doi:10.7554/eLife.23326.

Kakizawa, S., Kaneko, T., Hasegawa, S., and Hirano, T. (1993). Activation of somatolactin cells in the pituitary of the rainbow trout Oncorhynchus mykiss by low environmental calcium. *Gen. Comp. Endocrinol.* 91, 298–306. doi:10.1006/gcen.1993.1130.

Kakizawa, S., Kaneko, T., and Hirano, T. (1997). Effects of hypothalamic factors on somatolactin secretion from the organ-cultured pituitary of rainbow trout. *Gen. Comp. Endocrinol.* 105, 71–78. doi:10.1006/gcen.1996.6799.

Kakizawa, S., Kaneko, T., and Hirano, T. (1996). Elevation of plasma somatolactin concentrations during acidosis in rainbow trout (Oncorhynchus mykiss). *J. Exp. Biol.* 199, 1043–1051.

Kaneko, K. J., Gélinas, C., and Gorski, J. (1993). Activation of the silent progesterone receptor gene by ectopic expression of estrogen receptors in a rat fibroblast cell line. *Biochemistry* 32, 8348–8359.

Kaneko, T. (1996). Cell biology of somatolactin. *Int Rev Cytol* 169, 1–24. doi:10.1016/S0074-7696(08)61983-X.

Kausel, G., Salazar, M., Castro, L., Vera, T., Romero, A., Muller, M., and Figueroa, J. (2006). Modular changes of cis-regulatory elements from two functional Pit1 genes in the duplicated genome of Cyprinus carpio. *J. Cell Biochem.* 99, 905–921. doi:10.1002/jcb.20987.

Kausel, G., Vera, M.I., Figueroa,J.,Hernandez, J., San Martin, R., Molina, A., Chavez, V.M., Muller, M., Martial, J., Krauskopf, M. (1998). Effect of seasonal acclimatization on the expression of the carp transcription factor Pit-1. *Biochem. Mol. Biol. Int.* 45, 813-821.

Kausel, G., Vera, M. I., San Martin, R., Figueroa, J., Molina, A., Muller, M., Martial, J., and Krauskopf, M. (1999). Transcription factor Pit-1 expression is modulated upon seasonal acclimatization of eurythermal ectotherms: Identification of two Pit-1 genes in the carp. *J. Cell Biochem.*

Kausel, G., Vera, T., Valenzuela, G., Lopez, M., Romero, A., Muller, M., Figueroa, J. (2010). At least two expressed genes for transcription factors Pitx2 and Rpx are present in common carp and are upregulated during winter acclimatization. *Gen. Comp. Endocrinol.* 169, 250-257.

Kawauchi, H., Sower, S. A., and Moriyama, S. (2009). "Chapter 5 the neuroendocrine regulation of prolactin and somatolactin secretion in fish," in *Fish Neuroendocrinology* Fish Physiology. (Elsevier), 197–234. doi:10.1016/S1546-5098(09)28005-8.

Keller, T. E., and Yi, S. V. (2014). DNA methylation and evolution of duplicate genes. *Proc. Natl. Acad. Sci. USA* 111, 5932–5937. doi:10.1073/pnas.1321420111.

Kidd, K. A., Blanchfield, P. J., Mills, K. H., Palace, V. P., Evans, R. E., Lazorchak, J. M., and Flick, R. W. (2007). Collapse of a fish population after exposure to a synthetic estrogen. *Proc. Natl. Acad. Sci. USA* 104, 8897–8901. doi:10.1073/pnas.0609568104.

Kitano, H. (2004). Biological robustness. *Nat. Rev. Genet.* 5, 826–837. doi:10.1038/nrg1471.

Klinge, C. M. (2001). Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.* 29, 2905–2919. doi:10.1093/nar/29.14.2905.

Koganti, P. P., Wang, J., Cleveland, B., and Yao, J. (2017). 17β-Estradiol Increases Non-CpG Methylation in Exon 1 of the Rainbow Trout (Oncorhynchus mykiss) MyoD Gene. *Mar. Biotechnol.* 19, 321–327. doi:10.1007/s10126-017-9756-6.

Krinner, S., Heitzer, A. P., Diermeier, S. D., Obermeier, I., Längst, G., and Wagner, R. (2014). CpG domains downstream of TSSs promote high levels of gene expression. *Nucleic Acids Res.* 42, 3551–3564. doi:10.1093/nar/gkt1358.

Kronenberg, H., and Williams, R. H. (2008). *Williams textbook of endocrinology*. 11th ed. /. Philadelphia: Saunders/Elsevier.

Kuramoto, N., Baba, K., Gion, K., Sugiyama, C., Taniura, H., and Yoneda, Y. (2003). Xenobiotic response element binding enriched in both nuclear and microsomal fractions of rat cerebellum. *J. Neurochem.* 85, 264–273.

Lange, A., Paull, G. C., Hamilton, P. B., Iguchi, T., and Tyler, C. R. (2011). Implications of persistent exposure to treated wastewater effluent for breeding in wild roach (Rutilus rutilus) populations. *Environ. Sci. Technol.* 45, 1673–1679. doi:10.1021/es103232q.

Länge, R., Hutchinson, T. H., Croudace, C. P., Siegmund, F., Schweinfurth, H., Hampe, P., Panter, G. H., and Sumpter, J. P. (2001). Effects of the synthetic estrogen 17 alpha-ethinylestradiol on the life-cycle of the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem.* 20, 1216–1227.

Larhammar, D., and Risinger, C. (1994). Molecular genetic aspects of tetraploidy in the common carp Cyprinus carpio. *Mol. Phylogenet. Evol.* 3, 59–68. doi:10.1006/mpev.1994.1007.

Leppek, K., Das, R.,Barna,M. (2018). Functional 5'-URT mRNA structure in eukaryotic translation regulation and how to find them. Nature Rev. Mol. Cell. Biol. 19, 158-174.

Li, S., Crenshaw, E. B., Rawson, E. J., Simmons, D. M., Swanson, L. W., and Rosenfeld, M. G. (1990). Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* 347, 528–533. doi:10.1038/347528a0.

Li, Y., Ye, C., Shi, P., Zou, X.-J., Xiao, R., Gong, Y.-Y., and Zhang, Y.-P. (2005). Independent origin of the growth hormone gene family in New

World monkeys and Old World monkeys/hominoids. *J. Mol. Endocrinol.* 35, 399–409. doi:10.1677/jme.1.01778.

Liu, X.-H., Xie, B.-W., Wang, Z.-J., and Zhang, Y.-G. (2018). Characterization and expression analyses of somatolactin-α and -β genes in rare minnows (Gobiocypris rarus) following waterborne cadmium exposure. *Fish Physiol Biochem* 44, 983–995. doi:10.1007/s10695-018-0487-z.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262.

Long, H.K., Prescott, S.L., Wysocka, J. (2016). Ever-changing Landscapes: transcriptional enhancers in development and evolution. Cell 167, 1170-1187.

López, M., Figueroa, J., Kausel, G., Vera, M. I., and Krauskopf, M. (2001). In situ hybridization of somatolactin transcripts in the pituitary glands from acclimatized carp (Cyprinus carpio). *Biol. Res.* 34, 7–13. doi:10.4067/S0716-97602001000100010.

Lopez, M., Nica, G., Motte, P., Martial, J. A., Hammerschmidt, M., and Muller, M. (2006). Expression of the somatolactin beta gene during zebrafish embryonic development. *Gene Expr Patterns* 6, 156–161. doi:10.1016/j.modgep.2005.06.010.

Louis, E. J. (2007). Evolutionary genetics: making the most of redundancy. *Nature* 449, 673–674. doi:10.1038/449673a.

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi:10.1126/science.290.5494.1151.

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473.

Lynn, S. G., and Shepherd, B. S. (2007). Molecular characterization and sex-specific tissue expression of prolactin, somatolactin and insulin-like

growth factor-I in yellow perch (Perca flavescens). *Comp Biochem Physiol B, Biochem Mol Biol* 147, 412–427. doi:10.1016/j.cbpb.2007.02.005.

Masser, D. R., Stanford, D. R., and Freeman, W. M. (2015). Targeted DNA methylation analysis by next-generation sequencing. *J. Vis. Exp.* doi:10.3791/52488.

May, D., Todd, C. M., and Rand-Weaver, M. (1997). cDNA cloning of eel (Anguilla anguilla) somatolactin. *Gene* 188, 63–67. doi:10.1016/S0378-1119(96)00777-9.

Mayr, C. (2018). What are 3'UTRs doing? Cold Spring Harbor Perspectives in Biology, doi: Mayr, C. (2018). What Are 3' UTRs Doing? Cold Spring Harbor Perspectives in Biology, doi:10.1101/cshperspect.a034728

McGaughey, D. M., Abaan, H. O., Miller, R. M., Kropp, P. A., and Brody, L. C. (2014). Genomics of CpG methylation in developing and developed zebrafish. *G3 (Bethesda)* 4, 861–869. doi:10.1534/g3.113.009514.

McKenna, N. J., and O'Malley, B. W. (2000). An issue of tissues: divining the split personalities of selective estrogen receptor modulators. *Nat. Med.* 6, 960–962. doi:10.1038/79637.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol.* 3, REVIEWS0004.

Mingarro, M., Vega-Rubín de Celis, S., Astola, A., Pendón, C., Valdivia, M. M., and Pérez-Sánchez, J. (2002). Endocrine mediators of seasonal growth in gilthead sea bream (Sparus aurata): the growth hormone and somatolactin paradigm. *Gen. Comp. Endocrinol.* 128, 102–111.

Montefusco-Siegmund, R. A., Romero, A., Kausel, G., Muller, M., Fujimoto, M., and Figueroa, J. (2006). Cloning of the prepro C-RFa gene and brain localization of the active peptide in Salmo salar. *Cell Tissue Res.* 325, 277–285. doi:10.1007/s00441-006-0168-6.

Murdoch, F. E., Byrne, L. M., Ariazi, E. A., Furlow, J. D., Meier, D. A., and Gorski, J. (1995). Estrogen receptor binding to DNA: affinity for

nonpalindromic elements from the rat prolactin gene. *Biochemistry* 34, 9144–9150.

Nash, J. P., Kime, D. E., Van der Ven, L. T. M., Wester, P. W., Brion, F., Maack, G., Stahlschmidt-Allner, P., and Tyler, C. R. (2004). Long-term exposure to environmental concentrations of the pharmaceutical ethynylestradiol causes reproductive failure in fish. *Environ. Health Perspect.* 112, 1725–1733.

Németh, A., and Längst, G. (2004). Chromatin higher order structure: opening up chromatin for transcription. *Brief Funct Genomic Proteomic* 2, 334–343.

Nussey, S., and Whitehead, S. (2001). *Endocrinology: an integrated approach.* Oxford: BIOS Scientific Publishers.

Ocampo Daza, D., and Larhammar, D. (2018). Evolution of the growth hormone, prolactin, prolactin 2 and somatolactin family. *Gen. Comp. Endocrinol.* 264, 94–112. doi:10.1016/j.ygcen.2018.01.007.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* 120, 621–623.

Ohno, S. (1973). Ancient linkage groups and frozen accidents. *Nature* 244, 259–262. doi:10.1038/244259a0.

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin, Heidelberg: Springer Berlin Heidelberg doi:10.1007/978-3-642-86659-3.

Omori, Y. Kon, T. (2018). Goldfish: an old and new model system to study vertebrate development, evolution and human disease. J Biochem 2018 Sep 14, doi:10.1093/jb/mvy076.

Ono, M., Harigai, T., Kaneko, T., Sato, Y., Ihara, S., and Kawauchi, H. (1994). Pit-1/GH factor-1 involvement in the gene expression of somatolactin. *Mol. Endocrinol.* 8, 109–115. doi:10.1210/mend.8.1.8152425.

Ono, M., Takayama, Y., Rand-Weaver, M., Sakata, S., Yasunaga, T., Noso, T., and Kawauchi, H. (1990). cDNA cloning of somatolactin, a pituitary

protein related to growth hormone and prolactin. *Proc. Natl. Acad. Sci. USA* 87, 4330–4334.

Orgel, L. E. (1977). Gene-duplication and the origin of proteins with novel functions. *J. Theor. Biol.* 67, 773.

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi:10.1146/annurev.genet.34.1.401.

Páez-Pereda, M., Kovalovsky, D., Hopfner, U., Theodoropoulou, M., Pagotto, U., Uhl, E., Losa, M., Stalla, J., Grübler, Y., Missale, C., et al. (2001). Retinoic acid prevents experimental Cushing syndrome. *J. Clin. Invest.* 108, 1123–1131. doi:10.1172/JCI11098.

Pawlowski, S., van Aerle, R., Tyler, C. R., and Braunbeck, T. (2004). Effects of 17alpha-ethinylestradiol in a fathead minnow (Pimephales promelas) gonadal recrudescence assay. *Ecotoxicol. Environ. Saf.* 57, 330–345. doi:10.1016/j.ecoenv.2003.07.019.

Pendón, C., Martínez-Barberá, J. P., and Valdivia, M. M. (1994). Cloning of a somatolactin-encoding cDNA from sole (Solea senegalensis). *Gene* 147, 227–230.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45. doi:10.1093/nar/29.9.e45.

Pfäffle, R. W., DiMattia, G. E., Parks, J. S., Brown, M. R., Wit, J. M., Jansen, M., Van der Nat, H., Van den Brande, J. L., Rosenfeld, M. G., and Ingraham, H. A. (1992). Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. *Science* 257, 1118–1121.

Pham, T.-H., Minderjahn, J., Schmidl, C., Hoffmeister, H., Schmidhofer, S., Chen, W., Längst, G., Benner, C., and Rehli, M. (2013). Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res.* 41, 6391–6402. doi:10.1093/nar/gkt355.

Planas, J. V., Swanson, P., Rand-Weaver, M., and Dickhoff, W. W. (1992). Somatolactin stimulates in vitro gonadal steroidogenesis in coho salmon, Oncorhynchus kisutch. *Gen. Comp. Endocrinol.* 87, 1–5.

Pocock, T., and Falk, S. (2014). Negative impact on growth and photosynthesis in the green alga Chlamydomonas reinhardtii in the presence of the estrogen 17α-ethynylestradiol. *PLoS One* 9, e109289. doi:10.1371/journal.pone.0109289.

Pogoda, H.M., Hammerschmidt, M. (2007).Molecular genetics of pituitary development in zebrafish. *Semin. Cell Dev. Biol.* 18, 543-558.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., and Yan, Y.-L. (2004). Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* 20, 481–490. doi:10.1016/j.tig.2004.08.001.

Prince, V. E., and Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* 3, 827–837. doi:10.1038/nrg928.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. USA* 97, 5237–5242. doi:10.1073/pnas.97.10.5237.

Rand-Weaver, M., Noso, T., Muramoto, K., and 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. doi:10.1021/bi00220a010.

Rand-Weaver, M., Pottinger, T. G., and Sumpter, J. P. (1993). Plasma somatolactin concentrations in salmonid fish are elevated by stress. *J. Endocrinol.* 138, 509–515.

Rand-Weaver, M., Pottinger, T. G., and Sumpter, J. P. (1995). Pronounced seasonal rhythms in plasma somatolactin levels in rainbow trout. *J. Endocrinol.* 146, 113–119.

Rand-Weaver, M., and Swanson, P. (1993). Plasma somatolactin levels in coho salmon (Oncorhynchus kisutch) during smoltification and sexual maturation. *Fish Physiol Biochem* 11, 175–182. doi:10.1007/BF00004564.

Rand-Weaver, M., Swanson, P., Kawauchi, H., and Dickhoff, W. W. (1992). Somatolactin, a novel pituitary protein: purification and plasma levels during reproductive maturation of coho salmon. *J. Endocrinol.* 133, 393–403.

Rapp, R. A., and Wendel, J. F. (2005). Epigenetics and plant evolution. *New Phytol.* 168, 81–91. doi:10.1111/j.1469-8137.2005.01491.x.

Ravi, V., Venkatesh, B. (2018) The divergent genomes of teleosts. *Ann. Rev. Animal Biosci.* 6, 47-68.

Reyhanian, N., Volkova, K., Hallgren, S., Bollner, T., Olsson, P.-E., Olsén, H., and Hällström, I. P. (2011). 17α-Ethinyl estradiol affects anxiety and shoaling behavior in adult male zebra fish (Danio rerio). *Aquat. Toxicol.* 105, 41–48. doi:10.1016/j.aquatox.2011.05.009.

Rodin, S. N., Parkhomchuk, D. V., and Riggs, A. D. (2005). Epigenetic changes and repositioning determine the evolutionary fate of duplicated genes. *Biochemistry. (Mosc)* 70, 559–567.

Rodin, S. N., and Riggs, A. D. (2003). Epigenetic silencing may aid evolution by gene duplication. *J. Mol. Evol.* 56, 718–729. doi:10.1007/s00239-002-2446-6.

Rosenfeld, M.G., Briata, P., Dasen, J., Gleiberman, A.S., Kioussi, C., Lin, C., O'Connell, S.M., Ryan, A., Szeto, D.P., Treier, M. (2000). Multistep signaling and transcriptional requirements for pituitary organogenesis in vivo. *Recent Prog. Horm. Res.* 55, 1-12.

Ryan, A. K., and Rosenfeld, M. G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.* 11, 1207–1225.

Salla, R. F., Gamero, F. U., Rissoli, R. Z., Dal-Medico, S. E., Castanho, L. M., Carvalho, C. dos S., Silva-Zacarin, E. C. M., Kalinin, A. L., Abdalla, F.

C., and Costa, M. J. (2016). Impact of an environmental relevant concentration of 17α-ethinylestradiol on the cardiac function of bullfrog tadpoles. *Chemosphere* 144, 1862–1868. doi:10.1016/j.chemosphere.2015.10.042.

Seoighe, C. (2003). Turning the clock back on ancient genome duplication. *Curr. Opin. Genet. Dev.* 13, 636–643.

Simonet, N. G., Reyes, M., Nardocci, G., Molina, A., and Alvarez, M. (2013). Epigenetic regulation of the ribosomal cistron seasonally modulates enrichment of H2A.Z and H2A.Zub in response to different environmental inputs in carp (Cyprinus carpio). *Epigenetics Chromatin* 6, 22. doi:10.1186/1756-8935-6-22.

Sinha, Y. N. (1995). Structural variants of prolactin: occurrence and physiological significance. *Endocr. Rev.* 16, 354–369. doi:10.1210/edrv-16-3-354.

Sirotkin, A. V., and Harrath, A. H. (2014). Phytoestrogens and their effects. *Eur. J. Pharmacol.* 741, 230–236. doi:10.1016/j.ejphar.2014.07.057.

Smeets, J. M. W., Rankouhi, T. R., and Nichols, K. M. (1999). In VitroVitellogenin Production by Carp (Cyprinus carpio) Hepatocytes as a screening method for determining (anti) estrogenic activity of xenobiotics. *Toxicol Appl Pharmacol* 157, 68-76.

Solé, M., Porte, C., and Barceló, D. (2000). Vitellogenin induction and other biochemical responses in carp, Cyprinus carpio, after experimental injection with 17 alpha-ethynylestradiol. *Arch Environ Contam Toxicol* 38, 494–500.

Sonnenschein, C., and Soto, A. M. (1998). An updated review of environmental estrogen and androgen mimics and antagonists. *J. Steroid Biochem. Mol. Biol.* 65, 143–150.

Stefaneanu, L., Kovacs, K., Horvath, E., Lloyd, R. V., Buchfelder, M., Fahlbusch, R., and Smyth, H. (1994). In situ hybridization study of estrogen receptor messenger ribonucleic acid in human

adenohypophysial cells and pituitary adenomas. *J. Clin. Endocrinol. Metab.* 78, 83–88. doi:10.1210/jcem.78.1.8288720.

Suga, A., Nagae, M., and Yamaguchi, Y. (2018). Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation. *Glycobiology*. doi:10.1093/glycob/cwy059.

Sumpter, J. P., and Jobling, S. (2013). The occurrence, causes, and consequences of estrogens in the aquatic environment. *Environ. Toxicol. Chem.* 32, 249–251. doi:10.1002/etc.2084.

Takayama, Y., Rand-Weaver, M., Kawauchi, H., and Ono, M. (1991). Gene structure of chum salmon somatolactin, a presumed pituitary hormone of the growth hormone/prolactin family. *Mol. Endocrinol.* 5, 778–786. doi:10.1210/mend-5-6-778.

Taylor, J. S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615–643. doi:10.1146/annurev.genet.38.072902.092831.

Tsai, Z. T.-Y., Tsai, H.-K., Cheng, J.-H., Lin, C.-H., Tsai, Y.-F., and Wang, D. (2012). Evolution of cis-regulatory elements in yeast de novo and duplicated new genes. *BMC Genomics* 13, 717. doi:10.1186/1471-2164-13-717.

Uchida, K., Moriyama, S., Breves, J. P., Fox, B. K., Pierce, A. L., Borski, R. J., Hirano, T., and Grau, E. G. (2009). cDNA cloning and isolation of somatolactin in Mozambique tilapia and effects of seawater acclimation, confinement stress, and fasting on its pituitary expression. *Gen. Comp. Endocrinol.* 161, 162–170. doi:10.1016/j.ygcen.2008.11.031.

Valenzuela, G. E., Perez, A., Navarro, M., Romero, A., Figueroa, J., and Kausel, G. (2015). Differential response of two somatolactin genes to zinc or estrogen in pituitary of Cyprinus carpio. *Gen. Comp. Endocrinol.* 215, 98–105. doi:10.1016/j.ygcen.2014.09.015.

Van dePeer, Y., Taylor, J. S., and Meyer, A. (2003). Are all fishes ancient polyploids? *J. Struct. Funct. Genomics* 3, 65–73.

Vega-Rubín de Celis, S., Gómez, P., Calduch-Giner, J. A., Médale, F., and Pérez-Sánchez, J. (2003). Expression and characterization of European sea bass (Dicentrarchus labrax) somatolactin: assessment of in vivo metabolic effects. *Mar. Biotechnol.* 5, 92–101. doi:10.1007/s10126-002-0053-6.

Venkatesh, B. (2003). Evolution and diversity of fish genomes. *Curr. Opin. Genet. Dev.* 13, 588–592. doi:10.1016/j.gde.2003.09.001.

Wagner, A. (2001). Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends Genet.* 17, 237–239.

Walsh, J. B. (1995). How often do duplicated genes evolve new functions? *Genetics* 139, 421–428.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., et al. (2014). Genome sequence and genetic diversity of the common carp, Cyprinus carpio. *Nat. Genet.* 46, 1212–1219. doi:10.1038/ng.3098.

Yang, B.-Y., and Chen, T. T. (2003). Identification of a new growth hormone family protein, somatolactin-like protein, in the rainbow trout (Oncorhyncus mykiss) pituitary gland. *Endocrinology* 144, 850–857. doi:10.1210/en.2002-221005.

Zhang, H., Okamoto, N., and Ikeda, Y. (1995). Two c-myc genes from a tetraploid fish, the common carp (Cyprinus carpio). *Gene* 153, 231–236.

Zhao, Y., and Garcia, B. A. (2015). Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.* 7, a025064. doi:10.1101/cshperspect.a025064.

Zhu, Y., Stiller, J. W., Shaner, M. P., Baldini, A., Scemama, J.-L., and Capehart, A. A. (2004). Cloning of somatolactin alpha and beta cDNAs in zebrafish and phylogenetic analysis of two distinct somatolactin subtypes in fish. *J. Endocrinol.* 182, 509–518.

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

stressors or various backgrounds. *Gen. Comp. Endocrinol.* 99, 275–288. doi:10.1006/gcen.1995.1111.

Zhu, Y., Yoshiura, Y., Kikuchi, K., Aida, K., and Thomas, P. (1999). Cloning and phylogenetic relationship of red drum somatolactin cDNA and effects of light on pituitary somatolactin mRNA expression. *Gen. Comp. Endocrinol.* 113, 69–79. doi:10.1006/gcen.1998.7180.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481. doi:10.1038/nature12433.