



HAL
open science

Identification and structural characterization of the factors involved in vitellogenesis and its regulation in the African Osteoglossiforme of aquacultural interest *Heterotis niloticus* (Cuvier, 1829)

N'Zi Daniel Koua, Jesús Núñez-Rodríguez, Julie Orjuela, Céline Zatylny-Gaudin, Marie-Pierre Dubos, Benoît Bernay, Julien Pontin, Erwan Corre, Joël Henry

► To cite this version:

N'Zi Daniel Koua, Jesús Núñez-Rodríguez, Julie Orjuela, Céline Zatylny-Gaudin, Marie-Pierre Dubos, et al.. Identification and structural characterization of the factors involved in vitellogenesis and its regulation in the African Osteoglossiforme of aquacultural interest *Heterotis niloticus* (Cuvier, 1829). *General and Comparative Endocrinology*, 2020, pp.113532. 10.1016/j.ygcen.2020.113532 . hal-02867550

HAL Id: hal-02867550

<https://hal.science/hal-02867550>

Submitted on 15 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Identification and structural characterization of the factors involved in vitellogenesis and**
2 **its regulation in the African Osteoglossiforme of aquacultural interest *Heterotis niloticus***
3 **(Cuvier, 1829)**

4
5 N’Zi Daniel Koua^{abc}, Jésus Núñez-Rodríguez^d, Julie Orjuela^e, Céline Zatylny-Gaudin^{ac}, Marie-
6 Pierre Dubos^{ac}, Benoît Bernay^f, Julien Pontin^f, Erwan Corre^g et Joël Henry^{ac}

7
8 a: NORMANDIE UNIV, UNICAEN, CNRS, BOREA, 14000 CAEN, France
9 b: INP-HB, Département FOREN, BP 1313 Yamoussoukro, Côte d’Ivoire
10 c: Laboratoire de Biologie des Organismes et Ecosystèmes Aquatiques (BOREA) Université de
11 Caen-Normandie, MNHN, SU, UA, CNRS, IRD, Esplanade de la paix, 14032 Caen Cedex,
12 France
13 d: IRD UMR207, UAGRM, SANTA CRUZ, BOLIVIA.
14 e: IRD UMR BOREA, France.
15 f: NORMANDIE UNIV, UNICAEN, SF ICORE, Proteogen Platform, Esplanade de la paix, 14032
16 Caen, France
17 g: Sorbonne Université, CNRS, FR2424, ABiMS, Station Biologique, F-29680 Roscoff, France
18

19 **ABSTRACT**

20 The African bonytongue (*Heterotis niloticus*) is an excellent candidate for fish farming because it
21 has outstanding biological characteristics and zootechnical performances. However, the absence
22 of sexual dimorphism does not favor its reproduction in captivity or the understanding of its
23 reproductive behavior. Moreover, no molecular data related to its reproduction is yet available.
24 This study therefore focuses on the structural identification of the different molecular actors of
25 vitellogenesis expressed in the pituitary gland, the liver and the ovary of *H. niloticus*. A
26 transcriptomic approach based on *de novo* RNA sequencing of the pituitary gland, ovary and
27 liver of females in vitellogenesis led to the creation of three transcriptomes. *In silico* analysis of
28 these transcriptomes identified the sequences of pituitary hormones such as prolactin (PRL),
29 luteinizing hormone (LH) and follicle-stimulating hormone (FSH) and their ovarian receptors
30 (PRLR, FSHR, LHR). In the liver and ovary, estrogen receptors (ER) beta and gamma receptors,
31 liver vitellogenins (VtgB and VtgC) and their ovarian receptors (VLDLR) were identified.
32 Finally, the partial transcript of an ovarian Vtg weakly expressed compared to hepatic Vtg was
33 identified based on structural criteria. Moreover, a proteomic approach carried out from mucus
34 revealed the presence of one Vtg exclusively in females in vitellogenesis. In this teleost fish that
35 does not exhibit sexual dimorphism, mucus Vtg could be used as a sexing biomarker based on a
36 non-invasive technique compatible with the implementation of experimental protocols *in vivo*.

37
38 **Key words:** *Heterotis niloticus*, Reproduction, Vitellogenin, Neurohormones, Pituitary gland,
39 Estradiol receptors, Transcriptome, Proteome.

40

41 1. INTRODUCTION

42 In teleost fish as in all vertebrates, reproductive functions, from gametogenesis to sexual
43 behavior during pair formation, are mostly controlled by the hypothalamic-pituitary-gonadal axis
44 (Maruska & Fernald, 2011; Takahashi et al., 2016). Under the effect of sensory stimuli such as
45 environmental factors (Plant, 2015; Yang et al., 2017) and social factors (Maruska & Fernald,
46 2011), gonadotropin-releasing hormones (GnRH) produced by the brain stimulate the synthesis
47 of neurohormones such as prolactin (PRL), luteinizing hormone (LH) and follicle stimulating
48 hormone (FSH) by the pituitary gland (Zohar et al., 2010). PRL participates in the regulation of
49 several physiological processes including growth, reproduction and osmoregulation (Bu et al.,
50 2015; Freeman et al., 2000; Manzon, 2002; Whittington & Wilson, 2013). LH and FSH stimulate
51 the production of steroids that play a role in sexual differentiation by regulating gonad
52 development in both sexes, the reproductive cycle (gametogenesis, secondary sexual
53 characteristics, sexual behavior) and growth (Devlin & Nagahama, 2002; Taranger et al., 2010;
54 Zohar et al., 2010). These gonadotropins (GTHs) are used in aquaculture as inducers to improve
55 the reproduction of certain fish species (Mehdi & Ehsan, 2013; Mylonas et al., 2010). 17 β -
56 Estradiol (E2) is synthesized by the follicular cells of oocytes under the control of FSH (Hara et
57 al., 2016; Norris & Lopez, 2011); it is the main sex hormone that stimulates oogenesis and
58 vitellogenin (Vtg) synthesis by the liver in females (Hara et al., 2016; Erik R. Nelson & Habibi,
59 2013). This hormone is perceived *via* intracellular estrogen receptors (ERs) mainly expressed in
60 the liver (Nelson & Habibi, 2013; Unal et al., 2014).

61 Vitellogenin (Vtg) is a phospholipoglycoprotein precursor of vitelline proteins that accumulate in
62 female oocytes during vitellogenesis in the majority of oviparous species such as fish teleostans,
63 amphibians, reptiles, birds, most invertebrates, and platypus (Wallace et al., 1990; Zhang et al.,
64 2015). Vtg is mainly expressed in the liver, but also in non-hepatic tissues (Zhong et al., 2014)
65 where its level of expression is much lower (Wang et al., 2005). In general, teleost fish Vtg
66 comprises a heavy chain named lipovitelline (LvH or LvI), a phosphovitin domain rich in
67 phosphorylated serines (Pv), a light chain of lipovitelline (LvL or LvII), a β '-C component, and a
68 C-terminal coding region (CT) (Hara et al., 2016; Hiramatsu et al., 2006; Zhang et al., 2015) Vtg
69 plays a vital role in embryo development and larval survival. Methods based on the detection of
70 plasma Vtg in sexually mature adult individuals can be used to sex fish that do not display sexual
71 dimorphism.

72 For example, in *Arapaima gigas* that belongs to the order Osteoglossiformes and does not exhibit
73 sexual dimorphism (Chu-Koo et al., 2009; Dugué et al., 2008), the use of Vtg as a sexing marker
74 has made it possible to breed pairs and optimize reproduction in captivity (Carreiro, Furtado-

75 Neto, Mesquita, & Bezerra, 2011; Chu-Koo et al., 2009; IRD, 2008; Nuñez, 2008). Blood-borne
76 Vtg is incorporated into oocytes by endocytosis mediated by VtgR receptors integrated in the
77 oocyte membrane (Dominguez et al., 2014; Le Menn, 1979; Mañanós et al., 2007; Núñez-
78 Rodríguez et al., 1996; Stifani et al., 1990). VLDLRs are recycled following endocytosis (Hara
79 et al., 2016; Hiramatsu et al., 2006).

80 *Heterotis niloticus*, the species addressed in this study, is of economic interest because it is
81 important for inland fisheries and aquaculture in Africa. It shows exceptional growth
82 performances (Adite et al., 2006; Ezekiel & Abowei, 2013; Moreau, 1982; Odo et al., 2009).
83 Several technical barriers currently make it difficult to set up a zootechnical path compatible
84 with mass production. Biological particularities (Monentcham et al., 2009) such as the absence
85 of sexual dimorphism (Carreiro et al., 2011; Oladosu et al., 2007) observed in this
86 Osteoglossiforme species phylogenetically close to *A. gigas* (Betancur-R et al., 2017; Guo-Qing
87 & Wilson, 1996; Hilton, 2001, 2003; Lavoué, 2016; Lavoué & Sullivan, 2004; Nelson, 1968,
88 1969) makes it difficult to breed them in captivity or understand their reproductive behavior.
89 Information on Vtg (Chu-Koo et al., 2009; Dugué et al., 2008) is available about *A. gigas*, and
90 the structures of the α (Faria et al., 2013) and β (Sevilhano et al., 2017) subunits of LH and FSH
91 are known (Borella et al., 2009; Marcos & Adalberto, 2015), but no molecular data is yet
92 available on vitellogenesis and its regulation in *H. niloticus*.

93 The aim of this study was therefore to characterize Vtg(s) and their ovarian receptor(s), pituitary
94 neurohormones and their receptors, as well as E2 receptors in *H. niloticus* using a combination of
95 several "omic" approaches. A sexing method based on the presence of vitellogenin was further
96 tested from the mucus and blood of mature individuals using a proteomic approach.

97

98 **2. MATERIALS AND METHODS**

99 **2.1. Sampling and tissue samples**

100 The selected tissues were the pituitary gland, the ovary, the liver, the blood and the mucus of *H.*
101 *niloticus* from a fish farm located in central Côte d'Ivoire (5°08,713'W; 6°50,929'N) nearby the
102 political capital Yamoussoukro. The species was introduced for the first time in this country in
103 1957, from Cameroon (Lazard, 1990). Tissue samples were taken from fish anesthetized in water
104 containing 96° ethanol (1/10 dilution) (Bhanu & Philip, 2011). The samples were taken from
105 male and female individuals (whose sex was only known after dissection) during the
106 vitellogenesis period (May and June).

107 For RNA sequencing of the liver, ovary and pituitary gland, pooled samples from three females
108 were stored in RNAlater stabilization solution (Sigma) at 4°C. For proteomic analyses of blood

109 and mucus, the extraction medium was composed of 200 mM PBS, 0.1 mM EDTA, 1 mM DTT
110 and 400 mM PMSF, and the samples were stored at -80°C until extraction.

111
112

113 **2.2. Extraction of total RNAs**

114 The total RNAs of each tissue pool were extracted separately. The organs were ground in liquid
115 nitrogen, and 50 to 100 mg of tissue powder were mixed with 1 ml of TRizol (Ambion, Life
116 Technologies, Carlsbad, Calif., USA). Then extraction was performed according to the
117 manufacturer's recommendations. Total RNA was quantified using a NanoDrop
118 Spectrophotometer ND-1000 (NanoDrop Technologies, Inc.) and its integrity was assessed on a
119 2100 Bioanalyzer (Agilent Technologies). Libraries were generated from 250 ng of total RNA as
120 following: mRNA enrichment was performed using the NEBNext Poly(A) Magnetic Isolation
121 Module (New England BioLabs). cDNA synthesis was achieved with the NEBNext RNA First
122 Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New
123 England BioLabs). The remaining steps of library preparation were done using and the NEBNext
124 Ultra II DNA Library Prep Kit for Illumina (New England BioLabs). Adapters and PCR primers
125 were purchased from New England BioLabs. Libraries were quantified using the Quant-iT™
126 PicoGreen® dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised
127 Primers-SYBR Fast Universal kit (Kapa Biosystems). Average size fragment was determined
128 using a LabChip GX (PerkinElmer) instrument.

129

130 **2.3. RNA sequencing and *in silico* approach**

131 **Illumina sequencing, assembly and annotation**

132 The RNA samples were sequenced by the G enome Qu ebec platform
133 (<http://www.genomequebec.com/ressources-et-plateformes-technologiques.html>) on an Illumina
134 HiSeq4000 sequencer in 2 x 100 base pairs paired mode.

135 The raw readings were cleaned and filtered, and the traces of adapters were removed using
136 Trimmomatic v.0.33 (<http://www.usadellab.org/cms/index.php?page=trimmomatic>). The
137 readings were filtered using a quality threshold of 30 and a minimum size of 50 base pairs.
138 SortMeRNA v2.1 was used to filter ribosomal RNAs from reads (Kopylova et al., 2012). Finally,
139 the cleanup process was verified using fastQC v.0.11.5
140 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

141 The cleaned readings from the different libraries were subsequently assembled together using the
142 *de novo* transcriptome assembler Trinity v.2.5.1 (Grabherr et al., 2011). Relative abundances

143 were estimated using kallisto v.0.43.1 (Bray et al., 2016). To obtain the expression values
144 (FPKM: fragments per kilobase of exon per million fragments mapped) and identify low
145 coverage contigs (FPKM <1) and rare isoforms (<1%) so as to exclude them from the analysis,
146 the two software programs were launched *via* the Trinity package scripts. Peptide prediction was
147 performed using Transdecoder v.3.0.0 (Haas et al., 2013). A similarity search (blastx of
148 assembled transcripts and blastp of predicted peptides) was carried out against the uniprot-
149 swissprot database (version 09-2013). Signal peptides were predicted using signalP v4.0
150 (Petersen et al., 2011). Transmembrane domains were detected using TMHMM v2.0c (Krogh et
151 al., 2001). Finally, the functional annotation of the transcriptome was carried out using the
152 Trinotate v.3.0.1 pipeline (<http://trinotate.github.io>) described by Bryant et al. (2017).

153

154 ***In silico* analysis of the transcripts**

155 *In silico* analyses of the *H. niloticus* liver, ovary and pituitary transcripts resulting from the
156 assembly were mainly performed with Peptraq software developed internally (Zatylny-Gaudin et
157 al., 2016). Peptraq is a software program dedicated to the analysis of sequence files in fasta or txt
158 formats. It translates and filters transcripts and protein precursors on the basis of annotation
159 using keywords or structural criteria related to the presence of particular subsequences. Peptraq
160 can also assign the signal peptide, cleave it from preproprotein to proprotein, carry out the
161 cleavages of the convertases and identify the characteristic sequence repeats of the
162 neuropeptides.

163 The transmembrane domains were predicted by TMHMM v2.0
164 (www.cbs.dtu.dk/services/TMHMM). Secondary structures were obtained by alignment with
165 proteins of species whose domains were known or by searching for conserved domains on the
166 NCBI site (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb>. IRM). Tertiary structure was
167 modeled under I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>). The N-
168 glycosylation sites were identified under NetNGlyc 1.0 Server
169 (<http://www.cbs.dtu.dk/services/NetNGlyc/>). The nucleic and protein sequences were aligned
170 with Show translation (https://www.bioinformatics.org/sms/show_trans.html). Protein sequence
171 alignments of different species were performed under CLC Sequence Viewer 7.6.1
172 (<http://www.clcbio.com>). The amino acid identities across protein sequences were determined by
173 Protein Blast (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

174

175 **2.4. Proteomic approach focused on the search for Vtg**

176 **Sample preparation for MS analysis**

177 Proteomic analysis samples were first thawed in ice water and centrifuged for several minutes.
178 Proteins were precipitated at a rate of 20 μ l *per* 120 μ l of refrigerated acetone. The samples were
179 stored at -20°C overnight, and then centrifuged at 20,000 g for 20 minutes. The protein pellet
180 was resuspended in ammonium bicarbonate buffer (50 mM, pH 7). Proteins were digested
181 overnight at 37°C using 0.25 μ g of porcine trypsin (Promega, Madison, USA). The digestates
182 were desalted and concentrated on an Omix ZipTip μ C18 (Agilent) before analysis.

183

184 **NanoLC-MS/MS**

185 The chromatography step was performed on an ultra-high-pressure nano-chromatography system
186 (NanoElute, Bruker Daltonics). The peptides were concentrated on a C18 pepmap 100 (5 mm \times
187 300 μ m) precolumn (Thermo Scientific) and separated at 50 °C on a Reprosil reverse phase
188 column (25 cm \times 75 μ m, 1.6 μ m, C18) (Ionopticks). The mobile phases consisted of 0.1% formic
189 acid, 99.9% water (v/v) (A) and 0.1% formic acid in 99.9% ACN (v/v) (B). The flow rate was set
190 at 400 nl/min, and the gradient profile was as follows: from 2 to 15% B in 60 min, followed by a
191 25% increase in B over 30 min and then a 37% increase in 10 min, a 95% B wash step and a 2%
192 B rebalance.

193 MS analyses were performed on a TIMS-TOF mass spectrometer (Bruker Daltonics) with a
194 nano-electrospray ion source (CaptiveSpray, Bruker Daltonics). The system was calibrated
195 weekly, and the accuracy of the mass measurements was greater than 1 ppm. A capillary voltage
196 of 1,400 volts was used for ionization. The MS spectra were positive in the 100 to 1,700 m/z
197 mass range. The mass spectrometer was used in PASEF (parallel accumulation-serial
198 fragmentation) mode (Meier et al., 2015), excluding monocharged peptides. Ten PASEF MS/MS
199 scans were performed in 1.25 seconds from the 2-5 load range.

200

201 **Peptide sequencing and vitellogenin detection**

202 The fragmentation spectrum was used to determine peptide sequences. The database was
203 searched using Mascot 2.6.1 program (Matrix Science). Two local *H. niloticus* databases were
204 used: an ovarian protein database (437,476 entries) and a liver protein base (244,234 entries),
205 both constructed from the corresponding transcriptomes using Peptraq software.

206 The allowed variable modifications were as follows: C-carbamidomethyl, K-acetylation,
207 oxidation and dioxidation of methionine. "Trypsin" was selected with a tolerant mode including
208 two missing cleavage sites (Šlechtová et al., 2015). Mass accuracy was set at 20 ppm and 0.05
209 Da for the MS and MS/MS modes, respectively. The Mascot data was then transferred to Proline

210 validation software (<http://www.profi-proteomics.fr/proline/>) for data filtering with a significance
211 level of less than 0.05 and to eliminate redundant proteins.

212

213 **3. RESULTS**

214 **3.1. RNAseq and in silico approach**

215 A total of 1,003,989,936 raw reads paired with read lengths of 100 base pairs were generated.
216 After cleaning the poor quality adapters and sequences, 667,119,586 high quality paired readings
217 were used to generate a first overall assembly of 257,484 transcripts (corresponding to 156,636
218 Trinity "genes"). The length of the transcripts varies between 201 and 24,080 base pairs, an
219 average length of 1,008.5 base pairs and a median length of 418 base pairs. 90% of the cleaned
220 reads were successfully pseudo-aligned with the overall complete transcriptome, indicating
221 strong support for the transcriptome assembled by the reads. Weakly expressed transcripts
222 (FPKM <1) and rare isoforms (<1%) were excluded from the initial assembly, which resulted in
223 a filtered assembly of 62,852 transcripts (corresponding to 40,302 "genes" of Trinity), with
224 lengths between 201 and 18,966 base pairs, an average length of 1,382.9 base pairs and a median
225 length of 807 base pairs.

226

227 **Hepatic and ovarian Vitellogenin (Vtg)**

228 Three hepatic Vtgs and one ovarian Vtg have been identified in *H. niloticus*. Two of the Vtgs
229 expressed in the liver (HnVtg1 and HnVtg2) are represented by two transcripts for the N-
230 terminal end of 3,690 base pairs (N-ter1, FIG. 1a) and 3,669 base pairs (N-ter2, FIG. 1b) and by
231 a transcript in the C-terminal position of 1,518 base pairs (C-ter, FIG. 1c). The deduced protein
232 sequences show that N-ter1 (1,230 amino acids) and N-ter2 (1,223 amino acids) respectively
233 have 19 and 12 serines at the end of the sequences all represented by the same AGC codon and
234 corresponding to the N-terminal end of the phosvitin domain (FIG. 1a and 1b). The C-ter
235 transcript of 506 amino acids presents at the start of the sequence 20 serines also represented by
236 the AGC codon (FIG. 1c). The assembly of transcripts at the phosvitin domain is therefore
237 impracticable by Trinity since the codons corresponding to the polyserine site are identical.

238 One of the hepatic Vtgs (HnVtg3) has a complete 2,775 base pairs sequence encoding a 925
239 amino acids protein (FIG. 1d). As for ovary Vtg (HnVtg4), it came from a partial 930 base pairs
240 transcript encoding 310 amino acids (FIG. 1e). The secondary structure of *H. niloticus* Vtg
241 showed that hepatic Vtgs had a signal peptide of fifteen amino acids followed by only
242 lipovitellin domains (LvI and LvII) in HnVtg3 and by domains LvI, Pv, LvII, β '-C and CT in
243 HnVtg1 and HnVtg2 (FIG. 2). The incomplete ovary Vtg included the LvI and Pv domains.

244 Alignments showed that the divergence between N-ter1 and N-ter2 was at the beginning of the
245 sequence, with a more complete Pv domain in HnVtg1 (FIG. 3a). The Met¹ to Lys⁹⁰⁹ domains of
246 HnVtg1 and HnVtg3 (corresponding to the signal peptide and LvI of HnVtg3) were identical.
247 The Arg¹ domain at Ser²⁹⁷ of the incomplete ovarian Vtg (HnVtg4) was identical to the Arg⁸⁴⁴
248 domain at Ser¹¹⁴⁰ of N-ter1 and N-ter2 of hepatic Vtg; divergence occurred at the end of the
249 sequence in the Pv domain (FIG. 3b).

250 We successively aligned *Scleropages formosus* Vtg (1,745 amino acids, AWG47880) with *H.*
251 *niloticus* N-ter1 (FIG. 3c) and C-ter (FIG. 3d), and found only one orphan serine (FIG. 3e). The
252 homology between *H. niloticus* VtgC (HnVtg3) and VtgC from other teleost fish species was
253 analyzed; it showed 34.08% amino acid identity with *O. mykiss* (BBA57869) and 31.41% with
254 *O. niloticus* (XP_005459969) (FIG. 3f).

255

256

257 **Vitellogenin receptor**

258 Teleost vitellogenin receptors (VtgR) are classified as very low-density lipoprotein receptors
259 (VLDLR) belonging to the low-density lipoprotein superfamily (LDLR) (Li et al., 2003; Mizuta
260 et al., 2013; Pousis et al., 2012).

261 Two transcripts of 2,583 and 2,523 base pairs encoding two VLDLRs of 861 and 841 amino
262 acids were identified in the ovary (FIG. 4a) and the pituitary gland, respectively. HnVLDLR
263 comprised a signal peptide (Met¹-Arg²⁶) followed by seven and eight domains for the pituitary
264 and ovarian forms, respectively (FIG. 4b). HnVLDLR expressed in the ovary had an O-
265 glycosylated domain located between Ala⁷⁴⁸ and Ser⁷⁶⁷; this domain was not found in the
266 pituitary form. Apart from the O-glycosyl domain, the other domains of the two forms of
267 HnVLDLR had identical amino acid sequences.

268 The ligand-binding domain (LBD) was composed of eight ligand-binding repeats (LBRs) of 39
269 to 44 amino acids each, among which six cysteines and the SDE (LBR1-2, 4-7) and EDE
270 (LBR3) motifs. The precursor domains of the epidermal growth factor (EGF) was found in three
271 EGF1 (Cys³⁵⁵-Cys³⁸⁹), EGF2 (Cys³⁹⁵-Cys⁴²⁹) and EGF3 (Cys⁷⁰²-Cys⁷⁴⁵) repeats; each of them
272 comprised six cysteines.

273 Between EGF2 and EGF3 stood a domain comprising five 4-aa sequences of the YWS(V)D and
274 FWA(T)D types. A single 23-aa transmembrane domain (Ala⁷⁸⁶ to Trp⁸⁰⁸ for the ovarian form)
275 was present, and a cytoplasmic domain containing the FDNPVY sequence was found in C-
276 terminal position.

277 The analysis of amino acid sequence homology showed that apart from the O-glycosylated
278 domain that was absent from the VLDLRs of certain species, the particular sequences (SDE,
279 EDE, YWS(V)D, FWA(T)D and FDNPVY) were highly conserved. Moreover, HnVLDLR
280 shared a strong identity with VtgR or VLDLR of *S. formosus* (XP_018607358) (94%), *O.*
281 *bicirrhosum* (AXN72824) (92%), *O. mykiss* (CAD10640) (88%), *C. carpio* (XP_018935318)
282 (87%) and *O. aureus* (AAO27569) (85%) (FIG. 4c).

283

284 **17 β -estradiol receptors (ERs)**

285 The nomenclature chosen for the 17 β -estradiol receptors included the alpha subtype (ER α), the
286 beta or beta2 subtype (ER β) and the gamma or beta subtype 1 (ER γ).

287 Using *in silico* analysis, we identified the transcripts of four ER β and one ER γ : HnER β _X1
288 (1,674 base pairs) in the liver and the ovary (FIG. 5a), HnER β _X2 (1,653 base pairs) and
289 HnER β _X3 (1,281 base pairs) only in the liver (FIG. 5b), and HnER β _X4 (1,524 base pairs) and
290 HnER γ (1,443 base pairs) only in the ovary (FIG. 5c). No ER α was identified. Three receptors
291 linked to the estrogen receptors (ERRs) were also identified: HnERR α (1,296 base pairs) in the
292 liver, and HnERR γ _X1 (1,068 base pairs) and HnERR γ _X2 (1,101 base pairs) in the ovary (FIG.
293 6).

294 The deduced protein sequences of HnER β _X1, HnER β _X2, HnER β _X3, HnER β _X4, HnER γ ,
295 HnERR α , HnERR γ _X1 and HnERR γ _X2 comprised 558, 551, 427, 508, 481, 432, 356 and 367
296 amino acids, respectively.

297 The secondary structure of HnERs (FIG. 7) and HnERRs (FIG. 8) was composed of five
298 domains, i.e., i) the domain at the N-terminus (A/B) that contained the first AF-1 activation
299 function, ii) the DNA-binding domain (DBD), iii) flanking sequence D which promotes the
300 rotation of the DBD (Aranda & Pascual, 2001) and functions as a site of interaction with
301 corepressor proteins (Kumar et al., 2011), iv) the ligand-binding domain (LBD) where 17 β -
302 estradiol (E2) binds to amino acids to induce the second AF-2 activation function (Aranda &
303 Pascual, 2001; Kato et al., 1995), and v) a C-terminal region F whose specific function is as yet
304 unknown in fish (Aranda & Pascual, 2001; Nelson & Habibi, 2013).

305 The four forms of HnER β had identical DBD and D domains (FIG. 9a). The A/B domain of
306 HnER β _X3a had a reduced number (32) of little conserved amino acids at its N-terminal end,
307 whereas the other HnER β s had an identical 163-aa A/B domain. HnER β _X1 and HnER β _X3
308 had identical LBDs, whereas the LBDs of HnER β _X2 and HnER β _X4 did not contain the
309 sequence NMCVNSPE and the sequence located between Ser⁴⁹⁴ and Ala⁵²³ of HnER β _X1,
310 respectively. The F domain of HnER β _X4 contained 14 amino acids and differed from the other

311 HnERβs, which had an identical 35-aa F domain. HnERβ and HnERγ shared 58 to 68% amino
312 acid identity and had no common domain. HnERs shared less than 48% identity with HnERRs.
313 The different HnERRγ shared 78% identity with one another and less than 59% identity with
314 ERRα. However, the cores of the DBDs of the different receptors shared greater identity (FIG.
315 9b): 89% between ERβ and ERγ, 68 to 70% between ERs and ERRs, 97% between ERRγ_X1
316 and ERRγ_X2, and 94% between ERRγ and ERRα. The DBD core of all *H. niloticus* receptors
317 (ERs and ERRs) was composed of 66 amino acids including six cysteines followed by a COOH-
318 terminal extension (CTE) composed of 10 (HnERγ) to 25 (all ERRs) amino acids. The LBD was
319 much more variable than the DBD core: HnERβ and HnERγ shared 76 to 77% identity, while
320 HnERs shared less than 33% identity with HnERRs.

321 A comparison of *H. niloticus* ERβ with those of other teleost species showed that the core of the
322 DBD of HnERβ was identical to that of ERβ of *S. formosus* (XP_018583081.1) and *O.*
323 *bicirrhosum* (BAT68972. 1) and shared 98% identity with the EBβ DBD core of *O. mykiss*
324 (CAC06714), *O. aureus* (ACF75103) and *C. carpio* (BAF99814) (FIG. 9c). The LBD of
325 HnERβ_X1 shared 95% identity with *S. formosus* LBD, 93% with *O. bicirrhosum* LBD, and
326 82% with *O. mykiss*, *O. aureus* and *C. carpio* LBDs. A comparison of HnERγ DBD/LBD
327 sequences with the ERγs of other species showed 89/74% identity with *O. mykiss*
328 (NP_001118042), 89/72% with *O. aureus* (ACF75102.1), and 89/70% with *C. carpio*
329 (BAF99813.1) ERγ DBDs/LBDs (FIG. 9d). Amino acid identities across ERγs were lower than
330 across ERβs.

331 The P box involved in DNA recognition at the DBD level (Aranda & Pascual, 2001) was
332 represented by two peptides: EGCKA at the level of HnERs and EACKA at the level of
333 HnERRs. The D box involved in receptor dimerization (Aranda & Pascual, 2001) was
334 represented by four peptides: PATNQ (HnERβ), PANNQ (HnERγ), PASNE (HnERRα), and
335 PATNE (HnERRγ).

336 The six amino acids involved in the binding with E2 (Sabo-Attwood, Kroll, & Denslow, 2004)
337 were present in HnERβ_X1 (Glu³³⁰, Arg³⁷¹, Gly⁴⁹⁸, His⁵⁰¹, Leu⁵⁰² and Met⁵⁰⁵), HnERβ_X2
338 (Glu³³⁰, Arg³⁷¹, Gly⁴⁹⁸, His⁵⁰¹, Leu⁵⁰² and Met⁵⁰⁵), HnERβ_X3 (Glu³³⁰, Arg³⁷¹, Gly⁴⁹⁸, His⁵⁰¹,
339 Leu⁵⁰² and Met⁵⁰⁵) and HnERγ (Glu²⁷³, Arg³¹⁴, Gly⁴⁴¹, His⁴⁴⁴, Leu⁴⁴⁵ and Met⁴⁴⁸) (FIG. 5 and
340 FIG. 9). HnERβ_X2 had only two conserved amino acids involved in the binding with E2
341 (Glu³³⁰, Arg³⁷¹), which were absent in HnERRs. The LBDs of HnERβ_X1, HnERβ_X2,
342 HnERβ_X3 and HnERγ (FIG. 7) had a tertiary structure composed of 12 α-helices, with helix 12
343 located between Leu⁵¹³ and Leu⁵²¹, Leu⁵⁰⁶ and Leu⁵¹⁴, Leu³⁸² and Leu³⁹⁰, and Tyr⁴⁵⁷ and Asp⁴⁶⁵,

344 respectively. As for the LBD of HnER β _X2, it did not have helix 12, and its helix 11 was much
345 smaller than the helices 11 of the other HnER β s.

346

347 **Prolactin (PRL)**

348 Prolactin (PRL) is a multifunctional polypeptide neurohormone whose reproductive function in
349 fish includes the reproductive cycle and development, spawning, and parental care (Whittington
350 & Wilson, 2013). Expression of *H. niloticus* prolactin (HnPRL) in the pituitary gland was
351 identified from a 627 base pairs transcript encoding 209 amino acids (FIG. 10a).

352 The protein precursor of HnPRL had a 25-aa signal peptide (Met¹-Cys²⁵), and four major α -
353 helices representing 56% of the total protein, i.e., helix 1 (Phe²⁸-His⁵⁶), helix 2 (Glu⁶⁹-Ser¹¹⁵),
354 helix 3 (Gln¹²²-Lys¹⁴⁹) and helix 4 (Asp¹⁶⁹-Val²⁰¹), as well as two short α -helices between helix 1
355 and helix 2: helix 1' (Val⁶⁰-Lys⁶²) and helix 1'' (Lys⁸¹-Ser⁸⁶) (FIG. 10b). The tertiary structure of
356 HnPRL showed that the four major helices formed two antiparallel pairs, i.e., helix 1/helix 4 and
357 helix 2/helix 3 (FIG. 10c). The mature HnPRL contained four cysteines, one between helix 1'
358 and helix 1'' (Cys⁷⁰), two in helix 4 (Cys¹⁸² and Cys¹⁹⁹) and one at the end of the sequence
359 (Cys²⁰⁹). The alignments showed high or medium amino acid identity between HnPRL and the
360 PRLs of *S. formosus* (XP_018592416) (74%), *C. carpio* (CAA37063) (58%), *O. mykiss*
361 (AAA49611) (53%) and *O. niloticus* (AAA53281) (49%), but the four cysteines were conserved
362 (FIG. 10d).

363

364 **Prolactin receptor (PRLR)**

365 The HnPRL receptor (HnPRLR) was expressed in the ovary of females in vitellogenesis and was
366 represented by a 1,767 base pairs transcript encoding a 589-aa protein (FIG. 11a). HnPRLR
367 consisted of a 25-aa signal peptide (Met¹-Leu²⁵) followed by three domains (FIG. 11b), one 207-
368 aa extracellular domain (EC) (Ser²⁶-Arg²³²), one 23-aa transmembrane domain (TM) (Ser²³³-
369 Ile²⁵⁵), and one 334-aa intracellular domain (IC) (Asn²⁵⁶-Tyr⁵⁸⁹). Alignment of HnPRLR
370 sequences with the PRLRs of other teleost species showed 85% amino acid identity with *S.*
371 *formosus* (XP_018604730), 57% with *O. mykiss* (NP_001118071), 55% with *C. carpio*
372 (AAK95833), and 51% with *O. niloticus* (NP_001266477) (FIG. 11c).

373

374 **Alpha subunit of the glycoprotein hormone (GP α)**

375 An α -subunit of the glycoprotein hormone (GP α) common to the luteinizing hormone (LH) and
376 the follicle-stimulating hormone (FSH) (Acharjee et al.,2015) was characterized in the pituitary
377 gland of *H. niloticus*. The nucleic sequence of HnGP α comprised 345 base pairs encoding a 115-

378 aa protein (FIG. 12a). The secondary structure showed the presence of a 24-aa signal peptide and
379 three α -loop domains of 25, 28 and 23 amino acids, respectively (FIG. 12b). HnGP α had ten
380 cysteines (Cys¹ to Cys¹⁰) and α -loop 1 was located between Cys¹ and Cys⁴, α -loop 2 between
381 Cys⁵ and Cys⁶, and α -loop3 between Cys⁷ and Cys⁸ (FIG. 12a). Two N-glycosylation sites were
382 predicted, one in α -loop 2 (NITS) and the other in α -loop 3 (NHTD).

383 Sequence comparison showed that HnGP α was identical to *A. gigas* GP α (AIG51239) and that
384 some domains were common with *S. formosus* (XP_018588619.1) (82%), *C. carpio*
385 (CAA39852) (74%), *O. niloticus* (AAP49577) (65%) and *O. mykiss* (BAB17685) GP α (61%)
386 (FIG. 12c).

387

388 **β -subunit of the follicle-stimulating hormone (FSH β)**

389 The β -subunit transcript of *H. niloticus* follicle-stimulating hormone (HnFSH β) had a 381 base
390 pairs sequence encoding a 127-aa protein (FIG. 13a).

391 The secondary structure of HnFSH β was composed of a 19-aa signal peptide, three β -loop
392 domains 1, 2 and 3 of 26, 20 and 17 amino acids, respectively, and one 18-aa "seat-belt" loop
393 (FIG. 13b). The mature HnFSH β had 12 cysteines (Cys¹-Cys¹²). β -loop 1 was located between
394 Cys¹ and Cys⁴, β -loop 2 between Cys⁵ and Cys⁶, β -loop 3 between Cys⁷ and Cys⁸, and the "seat-
395 belt" loop between Cys¹⁰ and Cys¹² (FIG. 13a). A probable N-glycosylation site (NVSI) was
396 found in β -loop 1 between Cys¹ and Cys².

397 A comparison of the FSH β sequences showed that HnFSH β shared a high identity with FSH β of
398 *A. gigas* (AIA09918) (86%), but much less with *S. formosus* (KPP62167) (69%), *C. carpio*
399 (O13050) (51%), *O. mykiss* (BAB17686) (44%) and *O. niloticus* (AAP49575) (37%) FSH β s
400 (FIG. 13c).

401

402 **Follicle-stimulating hormone receptor (FSHR)**

403 The HnFSH receptor (HnFSHR) was expressed in the ovary of females in vitellogenesis and
404 corresponded to a 2,037 base pairs transcript encoding a 679-aa protein (FIG. 14a). HnFSHR had
405 a 37-aa signal peptide (Met¹ - Ala³⁷) followed by four extracellular domains (EC) of 322 (EC1),
406 21 (EC2), 19 (EC3) and 9 (EC4) amino acids, seven transmembrane domains of 23 amino acids
407 each, and four intracellular domains (IC) of 8 (IC1), 20 (IC2), 24 (IC3) and 58 (IC4) amino acids
408 (FIG. 14b). Alignment of HnFSHR sequences with FSHRs of other teleost species showed
409 64.94% amino acid identity with *O. mykiss* (NP_001117799) and 71.23% with *C. carpio*
410 (XP_018940722) (FIG. 14c).

411

412 **β -subunit of the luteinizing hormone (LH β)**

413 The β -subunit of *H. niloticus* luteinizing hormone (HnLH β) had a 423 base pairs sequence
414 encoding a 141-aa protein (FIG. 15a). The secondary structure made it possible to distinguish a
415 24-aa signal peptide, three β -loop domains 1, 2 and 3 of 26, 20 and 17 amino acids, respectively,
416 and one 18-aa seat-belt loop (FIG. 15b). Like HnFSH β , the mature HnLH β had twelve cysteines
417 (Cys¹-Cys¹²) and β -loop 1 was located between Cys¹ and Cys⁴, β -loop 2 between Cys⁵ and Cys⁶,
418 β -loop 3 between Cys⁷ and Cys⁸, and the “seat-belt loop” between Cys¹⁰ and Cys¹² (FIG. 15a). A
419 probable N-glycosylation site (NQTI) was located between Cys¹ and Cys². A comparison of LH β
420 sequences showed that HnLH β was identical to *A. gigas* LH β (LAO68014) and that it shared
421 domains with *O. mykiss* (BAB17687) (79%), *S. formosus* (KPP64307) (77%), *C. carpio*
422 (CAA42542) (74%) and *O. niloticus* (AAP49576) (68%) LH β s (FIG. 15c).

423

424 **Luteinizing hormone receptor (LHR)**

425 The HnLH receptor (HnLHR) was expressed in the ovary of females in vitellogenesis. It was a
426 2,124 base pairs transcript encoding a 708-aa protein (FIG. 16a). HnLHR had a 20-aa signal
427 peptide (Met¹-His²⁰) followed by four extracellular domains (EC) of 361 (EC1), 21 (EC2), 19
428 (EC3) and 9 (EC4) amino acids, seven transmembrane domains of 23 amino acids each, and four
429 intracellular domains (IC) of 8 (IC1), (IC2), 24 (IC3) and 65 (IC4) amino acids (FIG. 16b).

430 Alignment of the HnLHR sequences with the LHRs of other teleost species showed 80.37%
431 amino acid identity with *S. formosus* (XP_018593689), 65.23% with *O. mykiss* (NP_001117798)
432 and 62.77% with *O. niloticus* (XP_025753045) LHRs (FIG. 16c).

433

434 **3.2 Detection of Vtg in blood and mucus**

435 Mass spectrometry analysis detected the presence of Vtg in the blood and mucus of *H. niloticus*
436 females in vitellogenesis (FIG. 17). No Vtg or any of its cleavage products was detected in the
437 blood or mucus of males. Tryptic peptides of the LVI domain specific to HnVtg1 and HnVtg2
438 were detected in female blood samples, while only HnVtg1 peptides were detected in mucus
439 samples (FIG. 17a). Mapping tryptic peptides from the N and C-terminal HnVtg1 sequences
440 showed a coverage rate of 53.2% in blood proteins, and 31.9% in mucus proteins. As for the
441 different domains of Vtg, the coverage of the peptides detected in the blood was 42.5% for LVI,
442 0.8% for Pv and CT, 8.2% for LVII, and 1% for β '-C. In mucus, the coverage of the detected
443 peptides was 18.6% for LVI, 0.8% for Pv, 5.8% for LVII, 4.2% for β '-C, and 2.6% for CT (FIG.
444 17b).

445

446 **4. DISCUSSION**

447 Vtg and some of the molecular actors involved in vitellogenesis and its regulation were
448 characterized for the first time in the Osteoglossiforme *H. niloticus* using a combination of
449 "omic" approaches. Comparisons were made in most cases with *S. formosus*, a phylogenetically
450 close species, and with other phylogenetically more distant teleost species.

451
452 **PRL and its receptor**

453 Prolactin (PRL) is a polypeptide hormone secreted by the pituitary gland. Its importance in
454 reproductive physiology has been underlined in many teleost species such as *A. gigas* (Marcos &
455 Adalberto, 2015) and *Danio rerio* (Bu et al., 2015). Only one PRL expressed in the pituitary
456 gland of *H. niloticus* was identified. By contrast, several isoforms issued from alternative
457 splicing have been observed in mammals (Freeman et al., 2000) and in some non-mammalian
458 species including teleost fish that express two distinct forms of PRL (PRL1 and PRL2) (Huang et
459 al., 2009). As pointed out by Whittington & Wilson (2013), the expression of the PRL receptor
460 in *H. niloticus* ovary suggests the involvement of this neurohormone in vitellogenesis and
461 ovulation in this species. *H. niloticus* PRL (HnPRL) has a short 25-aa signal peptide like most
462 PRLs of teleost fish (Bu et al., 2015), while mammalian (Skorupski & Kmiec, 2013) and bird
463 (Bu et al., 2015) PRLs are made of up to 30 amino acids. The two cysteines in the N-terminal
464 position capable of forming a disulfide bridge in the PRLs of mammals (Bu et al., 2015; Teilum
465 et al., 2005), reptiles (Yasuda, Kawauchi, & Papkoff, 1990), Dipnoi (lungfish), chondrosteans
466 (sturgeons) and tetrapods (Noso et al., 1993) are absent from HnPRL as in most teleosts (Bu et
467 al., 2015; Watahiki et al., 1989). Although HnPRL shares variable amino acid identity (49 to
468 74%) with the PRLs of certain fish (FIG. 10d), it shares a common structure with the PRLs of
469 vertebrate species. It contains four major α -helices (helices 1 to 4) representing more than 50%
470 of the entire protein (Skorupski & Kmiec, 2013), and two short α -helices between helix 1 and
471 helix 2 forming a loop (Brooks, 2012; Teilum et al., 2005) that plays a vital role in protein
472 stabilization and ligand-receptor interaction (Goffin et al., 1996). Helix 1 and helix 4, on the one
473 hand, and helix 2 and helix 3, on the other hand, form two antiparallel pairs (Brooks, 2012;
474 Teilum et al., 2005). HnPRL contains four cysteines after propeller 1 that are well conserved
475 across species and are likely to form two disulfide bridges (Brooks, 2012; Bu et al., 2015;
476 Teilum et al., 2005). One is located between Cys⁷⁰ and Cys¹⁸², and the other at the C-terminal
477 level of the protein between Cys¹⁹⁹ and Cys²⁰⁹. Similarly to the PRL of its sister species *A. gigas*,
478 where immunocytochemistry showed that PRL had little affinity with Salmonid anti-PRL anti
479 serum (Borella et al., 2009), HnPRL shares little identity (53%) with *O. mykiss* PRL.

480

481 **Gonadotropins and their receptors**

482 *H. niloticus* FSH and LH expressed in the pituitary gland are heterodimeric glycoproteins like all
483 gonadotropic hormones (GTHs), with a common α -subunit called glycoprotein hormone α (GP α)
484 and specific β -subunits (Acharjee et al., 2015). A comparative analysis of the structure of *H.*
485 *niloticus* GTH with other members of the teleost group showed that the GP α subunit consisted of
486 three α loops and the β -subunits of three β loops with a "seat-belt" in the C terminal position.
487 HnGP α had the ten conserved cysteines believed to be involved in the formation of five disulfide
488 bridges, four proline residues, and two N-glycosylation sites at α loop 2 (NITS) and α loop 3
489 (NHTD). In catfish (*Heteropneust fossilis*), the N-glycosylation sites are located at loop 1 (NITS)
490 and loop 2 (NHTD) (Acharjee et al., 2015). Only eight cysteines are conserved in the GP α of
491 hagfish (Uchida et al., 2010). *A. gigas* GP α and LH β , described by Faria et al., (2013) and
492 Sevilhano et al., (2017), respectively, are identical to HnGP α and HnLH β identified in this study,
493 respectively. This result again demonstrates that there is a great phylogenetic proximity between
494 these two species of Osteoglossomorphs previously underlined by some authors (Betancur-R et
495 al., 2017; Guo-Qing & Wilson, 1996; Hilton, 2001, 2003; Lavoué, 2016; Lavoué & Sullivan,
496 2004; Nelson, 1968, 1969). This also completes the few studies reporting phylogenetic
497 hypotheses based on molecular data (Kumazawa & Nishida, 2000). The present study shows that
498 HnGP α shares between 61 and 82% amino acid identity with certain species of
499 Osteoglossiformes, Salmoniformes, Perciformes and Cypriniformes. *A. gigas* GP α is identical to
500 HnGP α and shares a higher amino acid identity (87 to 89.5%) with other teleost (Anguilliformes
501 and Ostariophysi) and Chondrosteans (Acipenseriformes) (Faria et al., 2013).

502 The position of the 12 cysteines and the only potential N-glycosylation site at LH β β -loop 1
503 (NXTX) observed in most vertebrates (Bousfield & Dias, 2011; So et al., 2005) is very well
504 conserved in *H. niloticus*. In catfish, the loop located between Cys¹⁰ and Cys¹¹ of the "seat-belt"
505 of LH β contains the receptor-binding sites of this hormone, whereas the loop corresponding to
506 Cys¹¹ and Cys¹² of LH β stimulates FSH receptor activity (Vischer et al., 2004).

507 In contrast to the LH β subunit, the FSH β subunit was 86% conserved between *H. niloticus* and
508 its sister species *A. gigas*, and diverged more with other species of Osteoglossiformes,
509 Perciformes, Salmoniformes and Cypriniformes (37 % to 69%). In *A. gigas*, the highest identity,
510 based on amino acid sequences, is 61% identity with Anguilliformes for FSH β and 76% with
511 Cypriniformes for LH β (Sevilhano et al., 2017). HnFSH β had 12 conserved cysteines; the first
512 four were located in β -loop 1 (Cys¹ to Cys⁴), Cys⁵ and Cys⁶ were located in β -loop 2, Cys⁷ and
513 Cys⁸ in β -loop 3, Cys⁹ between β -loop 3 and the "seat-belt loop", and Cys¹⁰ to Cys¹² in the "seat-

514 belt loop". In Cypriniformes and catfish such as *H. fossilus*, good conservation of the structure of
515 the FSH β subunit is observed, the only difference is that there is an additional cysteine between
516 the signal peptide and β -loop 1 in *H. fossilus*. Cys² is absent in Perciformes and so is Cys³ in
517 Salmonidae, but an additional cysteine residue in the N-terminal position, as in Cypriniformes
518 and catfish, gives 12 cysteines in total. In zebrafish, Cys¹⁰ and Cys¹¹, thought to be the decisive
519 loop of the FSH β "seat-belt" (Vischer et al., 2004) are absent, and two additional cysteines exist
520 in the N-terminal position before β -loop 1, allowing for a total number of 12 cysteines (So et al.,
521 2005). Given the importance of the "seat-belt" region for receptor interaction and heterodimer
522 formation, variation in the structure of FSH β subunits among fish species may result in
523 considerable differences in the nature of the receptor interactions and possibly the stability of the
524 heterodimer (Acharjee et al., 2015; Vischer et al., 2004). Whereas tetrapods, Dipnoi (lungfish),
525 Chondrosteans (sturgeons) and Holocephali (elephant sharks) have two probable sites of N-
526 glycosylation at the level of β -loop 1 of their FSH β subunit, in *H. niloticus*, only one N-
527 glycosylation site (NVSI) was observed, as in most teleosts. GTH glycosylation is essential for
528 bioactivity, disulfide bond formation, secretion rate, circulatory persistence, and signal
529 transduction (Ulloa-Aguirre et al., 1999). In humans, glycosylation site 1 has been associated to
530 signal transduction, while site 2 is linked to thermal stability, which has a greater effect on
531 disulfide bond formation and hormone secretion (Feng et al., 1995). The loss of the second N-
532 glycosylation of *H. niloticus* FSH β could therefore affect interactions among receptors
533 (Bousfield & Dias, 2011; Ulloa-Aguirre et al., 1999). Thus, the absence of a second
534 glycosylation site in the FSH β of most fish appears to strongly diverge from the highly
535 conserved vertebrate model, and raises the hypothesis of the incidence on the half-life of the
536 molecule and in turn on the absence of its detection in certain fish (Acharjee et al., 2015).

537

538 **Estradiol receptors**

539 Several authors agree on the existence of three distinct subtypes of 17 β -estradiol receptors (ERs)
540 in fish, with quite varied nomenclatures (Choi & Habibi, 2003; Hawkins et al., 2000; Hawkins &
541 Thomas, 2004; Marlatt et al., 2008; Métivier et al., 2002; Nelson & Habibi, 2013; Nelson &
542 Habibi, 2010; Norris & Lopez, 2011; Sabo-Attwood et al., 2004). These are the alpha (ER α ,
543 ESR1, ER, ESR, ESRA, Era or NR3A1), gamma or beta 1 (ER γ , ESR2 α , ER β -I, ER $\beta\alpha$, ER β 1)
544 and beta or beta 2 (ER β , ESR2b, ER β - II, ER β b, ER β 2 or NR3A2) receptors.

545 We identified the beta subtype of the estradiol receptor (ER β) in the liver of *H. niloticus* females
546 in vitellogenesis, and the beta and gamma subtypes (ER β and ER γ) in the ovary. The three ER
547 subtypes appeared to be differentially distributed in the tissues. The alpha subtype was not

548 present in any tissue, ER γ was expressed in the ovary, and ER β in the liver and ovary.
549 Investigations into the functional significance of the different ER subtypes in fish and other
550 vertebrates showed that expression patterns differed according to sex, gametogenesis stage,
551 species, and tissue type (Couse et al., 1997; Marlatt et al., 2008; Nelson et al., 2007) and that ER
552 subtypes bound to E2 with different levels of affinity (Genovese et al., 2014; Nelson & Habibi,
553 2013). For example, in *Micropterus salmoides* (Sabo-Attwood et al., 2004), the three ER
554 subtypes were detected in the liver, ovary and pituitary gland of females in early vitellogenesis,
555 with higher expression levels in the ovary. Of the three *M. salmoides* ER subtypes, ER β had the
556 highest level of expression in the ovary and liver. ER γ was more expressed in the ovary than
557 ER α , whereas the opposite was observed in the liver. Many studies have been devoted to the
558 study of the regulation of ERs in the liver (Nelson & Habibi, 2013) because Vtg, the precursor of
559 the proteins of the yolk of the egg, is mainly expressed in this tissue (Dominguez et al., 2012;
560 Hara et al., 2016; Prat et al., 1998). The strong expression of ER β and ER γ in the ovary of
561 certain species suggests that these two subtypes play a greater role in gene regulation in this
562 organ (Sabo-Attwood et al., 2004) through expression and stimulation of Vtg synthesis in some
563 fish (Kong et al., 2014; Wang et al., 2005; Wang et al., 2010).

564 Two types of estradiol receptor-related receptors (ERR α and ERR γ) were also identified in the
565 liver and ovary of *H. niloticus*. ERRs are orphan receptors for which no endogenous ligand has
566 yet been identified, so that these receptors have not received as much attention as some
567 endocrine receptors (Audet-Walsh & Giguère, 2015; Giguère, 1999). *H. niloticus* ERRs showed
568 significant homology with ERs, with 68-70% amino acid identity at the core of the DBD.
569 However, these ERRs do not bind to estrogens, but rather to hexanucleotide sequences in the
570 form of monomers or to composite elements in the form of dimers (Giguère, 1999). As a result,
571 ERRs do not participate directly in conventional estrogen signaling pathways or biological
572 processes (Audet-Walsh & Giguère, 2015). The two *H. niloticus* ER subtypes contained five
573 domains, as most nuclear receptors do (Aranda & Pascual, 2001; Nelson & Habibi, 2013; Sabo-
574 Attwood et al., 2004). The N-terminal region (A/B) contained the first activation function (AF-
575 1); the strong variation observed between HnER β and HnER γ (58 to 68%) and between
576 HnER β _X3 and the other HnER β s contributed to a high specificity of the receptors (Aranda &
577 Pascual, 2001). The core of the DBDs comprised 66 amino acids and was by far the most
578 conserved domain across HnERs (89%) and HnERRs (> 94%), as in most nuclear receptors of
579 different species (Aagaard et al., 2011). In *M. salmoides* (Sabo-Attwood et al., 2004) and
580 *Micropogonias undulatus* (Hawkins et al., 2000), the DBD nucleus of the β and γ forms of ERs
581 share around 70% identity, and therefore the protein sequence of ER γ is probably derived from

582 the duplication of ER β genes (Hawkins et al., 2000; Hawkins & Thomas, 2004). In addition to
583 the DBD core, the C-terminal extension (CTE), composed of the first 10-30 residues following
584 the fully conserved VGM sequence of the DBD nucleus, is an important determinant of specific
585 DNA binding for many nuclear receptors (Aagaard et al., 2011). Unlike the DBD nucleus, CTE
586 is more variable among nuclear receptors (Aagaard et al., 2011).

587 The amino acids considered to be essential for the recognition of estradiol in most vertebrate ERs
588 (Sabo-Attwood et al., 2004) were well preserved at the level of HnER β _X1, HnER β _X2,
589 HnER β _X3 and HnER γ , while HnER β _X4 possessed only two preserved sites.

590 The analysis of the predicted amino acid sequences revealed that the four forms of HnER β had
591 identical domains, and only diverged at the level of the A/B or LBD domains. These were four
592 HnER β isoforms resulting from alternative splicing of the gene coding for these estrogenic
593 receptors. In humans five isoforms of ER β have been identified and their divergence is located at
594 the C-terminus of the LBD and F domain (Leung et al., 2006). The human ER β 1 isoform has the
595 longest LBD sequence, containing the complete helices 11 and 12 where the second AF-2
596 activation function and four amino acids involved in binding with E2 are found; it is the only
597 functional isoform in homodimer. The C-terminal end of the LBD is much shorter in the other
598 isoforms because of the disorientation or absence of helices 11 and 12. As a result, they cannot
599 form functional homodimers but rather heterodimers with the ER β 1 isoform insofar as a single
600 functional helix 12 of the dimer is sufficient (Leung et al., 2006). Under these conditions,
601 HnER β _X1, HnER β _X2, HnER β _X3 and HnER γ , which possess helices 11 and 12 with all the
602 amino acids involved in the binding with E2, can function in homodimers, while HnER β _X4,
603 which does not have helix 12 and has a short helix 11 lacking the four amino acids involved in
604 the binding with E2, can only form functional heterodimers with the other HnER β s. The D-box
605 involved in receptor dimerization (Aranda & Pascual, 2001) is represented by different peptides,
606 namely PATNQ in HnER β and PANNQ isoforms in HnER γ , suggesting that these two types of
607 ERs do not form functional heterodimers on DNA (Zechel et al., 1994).

608

609 **Vitellogenins and their receptors**

610 Three forms of Vtg identified in many species (Schilling et al., 2015; Williams et al., 2014;
611 Yilmaz et al., 2015) are divided into two categories. One is characterized by the presence of all
612 domains, the other by the absence of one or more domains. In the first category, Vtg are
613 characterized by the presence of a phosphovitin domain rich in phosphorylated serines (Pv), a
614 lipovitellin heavy chain (LvI), a lipovitellin light chain (LvII), a β '-C component, and a C-
615 terminal coding region (CT). This category of Vtg has two variants, namely VtgA and VtgB,

616 which have more or less similar molecular weights. In the second category, VtgC lacks the Pv
617 domain (Hara et al., 2016; Hiramatsu et al., 2006; Yilmaz et al., 2015).

618 We identified four Vtg in *H. niloticus*. Many fish species have several Vtg. For example, three
619 Vtg were identified in *Larimichthys crocea* (Gao et al., 2019) and seven in *Danio rerio* (Wang et
620 al., 2005). The fish genome can contain up to 20 copies of genes encoding vitellogenins, e.g.,
621 *Oncorhynchus mykiss* (Trichet et al., 2000). HnVtg1, HnVtg2 and HnVtg4 have a Pv domain
622 which is used to classify them in the Vtg A or Vtg B group, whereas HnVtg3 corresponds to a
623 VtgC devoid of the Pv, β '-C and CT domains. Alignments made with the variants of Vtg A or B
624 of other fish species made it possible to classify HnVtg1 and HnVtg2 among VtgBs. *H. niloticus*
625 hepatic Vtg share high amino acid identity (> 98%) and appear to be derived from alternative
626 splicing of the gene encoding this protein. In *O. mykiss*, the genes encoding different Vtg show
627 98.7% similarity (Trichet et al., 2000) whereas *Morone saxatilis* VtgC shares less than 32%
628 identity with the A or B forms (Williams et al., 2014). The identification of Vtg transcripts in the
629 liver and ovary of *H. niloticus* revealed that Vtg are expressed in these two organs in this species.
630 Vtg is mainly expressed in the liver, but its expression has also been observed in the ovaries of
631 certain fish species such as *Danio rerio* (Wang et al., 2005), *Tanichthys albonubes* (Wang et al.,
632 2010) and *Rhodeus uyekii* (Kong et al., 2014). Ovarian Vtg, like most Vtg expressed in non-
633 hepatic organs or tissues, represent less than 10% of the level of hepatic Vtg expression (Wang et
634 al., 2005).

635 The N-ter1, N-ter2 and C-ter transcripts characteristic of HnVtg1 and HnVtg2 were not
636 sufficient to obtain the definitive sequences of these proteins because the exact number of serines
637 to be retained for the Pv domain was uncertain. The nineteen and twelve serines at the end of the
638 sequence of N-ter1 and N-ter2, respectively, as well as the twenty serines at the beginning of the
639 C-ter sequence all came from the same AGC codon, while serine is coded by five other codons
640 (TCT, TCC, TCA, TCG and AGT). Trinity failed to connect the N-ter and the C-ter. However,
641 alignment with the Vtg of *S. formosus*, another Osteoglossiforme, showed that the number of
642 serines in the Pv domain was uncertain; but this uncertainty was low and probably corresponded
643 to one or two serine residues; knowing that they may be missing or supernumerary serines, the
644 transcripts may be overlapping. This means that apart from a few serines, the HnVtg1 and
645 HnVtg2 sequences were complete. In other words, either some serines were missing to connect
646 the 2 N-ter with the C-ter, or the N-ter and C-ter overlapped. But as previously stated, serines in
647 the Pv domain of *Heterotis* Vtg are encoded by a single codon that makes assembly complex.

648 MS/MS analysis detected the presence of Vtg in the blood and mucus of *H. niloticus* females in
649 vitellogenesis. Enzyme immunoassays (ELISA) have been used in many species to detect the

650 presence of plasma Vtg (Bulukin et al., 2007; Chu-Koo et al., 2009; Ndiaye et al., 2006; Nuñez-
651 Rodriguez et al., 1989; Nunez Rodriguez et al., 1997; Roy et al., 2004) and mucus Vtg (Arukwe
652 & Røe, 2008; Genovese et al., 2014; Gordon et al., 1984; Meucci & Arukwe, 2005; Moncaut,
653 Nostro, & Maggese, 2003). The proteomic approach allowed us to sex *H. niloticus* adults
654 although the species is devoid of sexual dimorphism. In the absence of direct observation of
655 gonadal tissue, a similar sexing test was performed on blood samples from *Epinephelus*
656 *lanceolatus* individuals (Om et al., 2013). The fact that we did not detect Vtg in the blood and
657 mucus of *H. niloticus* males indicates that the sampled fish were not subject to endocrine
658 disruption due to mimetic substances of natural estrogens known as xenoestrogens contained in
659 certain foods distributed to most farmed fish (Pelissero & Sumpter, 1992; Pelissero et al., 1989)
660 or in the aquatic environment (Arukwe & Røe, 2008; Dugué et al., 2008; Genovese et al., 2014;
661 Le Menn, 1979).

662
663 Concerning Vtg receptors, two types of Vtg receptor transcripts apparently derived from
664 alternative splicing of the O-glycosylated domain (Chen et al., 2016) were identified in *H.*
665 *niloticus*. These were LR8 forms with eight LBRs in the LBD (Sappington & Raikhel, 1998) that
666 have long been considered the only VLDLR capable of binding Vtg in some oviparous species
667 (Reading et al., 2014). LR8+, which has an O-glycosylated domain, and LR8-, which does not
668 have one, were detected in the ovary and pituitary gland of *H. niloticus*, respectively. These two
669 forms of LR8 have also been identified in teleost fish such as *O. mykiss* (Prat et al., 1998) and
670 *Oncorhynchus clarki* (Mizuta et al., 2013). However, LR8- is expressed primarily in the ovary of
671 these species, while LR8+ is expressed in both the ovary and non-ovarian tissues such as the
672 brain, intestine and gills. Detection of LR8- in non-ovarian tissue (*H. niloticus* pituitary gland)
673 requires further investigation. In recent years, a new VLDLR called LRP13 or LR13+1 has been
674 identified in several teleost species other than *H. niloticus*, such as *M. saxatilis*, *Morone*
675 *americana* (Reading et al., 2014), *Cynoglossus semilaevis*, *Oryzias latipes* (Wang et al., 2017)
676 and *O. clarki* (Mushirobira et al., 2015), with thirteen LBRs in the N-terminal LBD and one LBR
677 in the C-terminal LBD (Hiramatsu et al., 2013). Some fish species may have up to four forms of
678 VLDLR, e.g., *M. Americana* (Reading et al., 2011). In the brain, VLDLRs play a role in
679 extracellular signaling, intracellular signaling processes and central nervous system development
680 (Trommsdorff et al., 1999).

681 LR8 are conserved structurally in many species (Dominguez et al., 2012; A. Li et al., 2003;
682 Pousis et al., 2012; Prat et al., 1998) including *H. niloticus*, which shares more than 84% amino
683 acid identity with species of Osteoglossiformes, Salmoniformes, Perciformes and Cypriniformes.

684 The precursors of HnVLDLR have a signal peptide followed by eight or seven domains for the
685 LR8+ and LR8- forms, respectively. LBD controls interactions between the receptor and the Vtg
686 Lv domain (Hiramatsu et al., 2001). It is composed of eight LBRs of around 40 amino acids
687 each, including six cysteines (Li et al., 2003; Luo et al., 2013) and negatively charged residues
688 on its surface. The EDE signature can be found in LBR3, and the SDE signature in other LBRs
689 well conserved across species (Agulleiro et al., 2007). Three LBRs (LBR1 to LBR3) are
690 primarily involved in the binding of Vtg to the receptor (Li et al., 2003). In chicken (Bajari et al.,
691 1998), the EDE tripeptide of LBR3 has a higher negative charge density than the SDE tripeptide
692 of other LBRs and appears to be determining for binding Vtg to the receptor, whereas a role in
693 calcium cage formation seems more likely in *O. niloticus* (Li et al., 2003). Cysteines from the
694 three precursor domains of the epidermal growth factor (EGF) are also well conserved across
695 species. The domain predicted to form a beta helical structure (Agulleiro et al., 2007) between
696 EGF2 and EGF3 is composed of five 4-aa sequences of the YWS(V)D and FWA(T)D types;
697 these are well conserved across species. Finally, the highly conserved FDNPVY sequence of the
698 cytoplasmic domain, potentially involved in receptor internalization after ligand binding
699 (Agulleiro et al., 2007) is also present in HnVLDLR.

700

701 **CONCLUSION/OUTLOOKS**

702 This study provides molecular data on vitellogenesis for the first time in an African
703 Osteoglossiforme named *Heterotis niloticus*. The nucleotide sequences and protein sequences
704 deduced from prolactin and its receptor, gonadotropins (FSH and LH) and their receptors, beta
705 and gamma receptors to estradiol and vitellogenins and their receptors have been identified. No
706 ER α has been identified, and the definitive sequences of two hepatic Vtg and ovarian Vtg remain
707 to be determined by supplementary analyses. We sexed the species by mass spectrometry
708 detection of Vtg in mucus and blood. This study can be considered as a necessary step in the
709 functional characterization of the identified molecules. It also makes it possible to envisage
710 studies of reproductive behavior from sexed individuals to better control reproduction and
711 breeding of this species.

712 Because MS and ELISA analyses are quantitative, it is urgent to establish a relationship between
713 the Vtg concentration in mucus and the maturity stage of females (Baumann et al., 2013;
714 Chatakondi & Kelly, 2013) so as to implement multiple applications in fish farming in non-
715 invasive conditions. This will preserve the welfare of farmed fish and thus produce quality,
716 relatively stress-free fish.

717 On a practical level, knowledge of the primary sequences of pituitary hormones will make it
718 possible to synthesize peptides to induce gametogenesis and control fry production continuously
719 throughout the year (Bry et al., 1978; Mehdi & Ehsan, 2013; Mylonas et al., 2010). This is an
720 important parameter for ensuring regular production of marketable fish.

721 Moreover, based on knowledge of *H. niloticus* pituitary neurohormones, it will be possible to
722 monitor the circulating levels in this species with specific quantitative tools such as ELISA or
723 mass spectrometry during *in vivo* experiments that remain to be undertaken. In the near future, it
724 will be necessary to characterize the whole neuropeptidome of this species and also the whole
725 regulatory peptidome to facilitate the interpretation of physiological studies and thus totally
726 master breeding. The use of molecular tools could facilitate the selection of successful breeders
727 (Hallerman, 2006) to develop sustainable, environment-friendly fish farming.

728

729

730 **AUTHOR INFORMATION**

731 **Corresponding author**

732 *E-mail: joel.henry@unicaen.fr Tel : +33 (0)2 31 56 55 96

733

734 **THANKS**

735 We thank Sabbat Affif, former owner of the fish farm where the samples were collected for this
736 study, and the farm manager Jacques Manlan who organized the various catches. The authors
737 also thank the IRD itrop HPC (Montpellier SouthGreen Platform: <https://bioinfo.ird.fr/>-
738 <http://www.southgreen.fr>) and the ABiMS platform of the Roscoff biological station [http://](http://abims.sb-roscoff.fr)
739 abims.sb-roscoff.fr) for providing the HPC resources that contributed to the search results
740 reported in this document.

741

742 **REFERENCES**

- 743 Aagaard, M. M., Siersbæk, R., & Mandrup, S. (2011). Biochimica et Biophysica Acta Molecular
744 basis for gene-specific transactivation by nuclear receptors ☆. *Biochimica et Biophysica*
745 *Acta*, 1812, 824–835. <https://doi.org/10.1016/j.bbadis.2010.12.018>
- 746 Acharjee, A., Chaube, R., & Joy, K. P. (2015). Molecular Cloning and Characterization of the
747 Gonadotropin Subunits GP α, FSH β, and LH β Genes in the Stinging Catfish
748 *Heteropneustes fossilis*: Phylogeny, Seasonal Expression, and Pituitary Localization.
749 *Journal of Experimental Zoology*, 323A, 567–585. <https://doi.org/10.1002/jez.1949>
- 750 Adite, A., Winemiller, K. O., & Fiogbe, E. D. (2006). Population structure and reproduction of

751 the African bonytongue *Heterotis niloticus* in the So River-floodplain system (West Africa):
752 implications for management. *Ecology of Freshwater Fish*, 15(1), 30–39.
753 <https://doi.org/10.1111/j.1600-0633.2005.00119.x>

754 Agulleiro, M. J., André, M., Morais, S., Cerdà, J., & Babin, P. J. (2007). High Transcript Level
755 of Fatty Acid-Binding Protein 11 but Not of Very Low-Density Lipoprotein Receptor Is
756 Correlated to Ovarian Follicle Atresia in a Teleost Fish (*Solea senegalensis*)¹. *Biology of*
757 *Reproduction*, 77(3), 504–516. <https://doi.org/10.1095/biolreprod.107.061598>

758 Aranda, A., & Pascual, A. (2001). Nuclear hormone receptors and gene expression.
759 *Physiological Reviews*, 81(3), 1269–1304. <https://doi.org/0031-9333/01>

760 Arukwe, A., & Røe, K. (2008). Molecular and cellular detection of expression of vitellogenin
761 and zona radiata protein in liver and skin of juvenile salmon (*Salmo salar*) exposed to
762 nonylphenol. *Cell and Tissue Research*, 331(3), 701–712. [https://doi.org/10.1007/s00441-](https://doi.org/10.1007/s00441-007-0543-y)
763 [007-0543-y](https://doi.org/10.1007/s00441-007-0543-y)

764 Audet-Walsh, É., & Giguère, V. (2015). The multiple universes of estrogen-related receptor α
765 and γ in metabolic control and related diseases. *Acta Pharmacologica Sinica*, 36, 51–61.
766 <https://doi.org/10.1038/aps.2014.121>

767 Bajari, T. M., Lindstedt, K. A., Riepl, M., Mirsky, V. M., Nimpf, J., Wolfbeis, O. S., ...
768 Schneider, W. J. (1998). A minimal binding domain of the low density lipoprotein receptor
769 family. *Biological Chemistry*, 379(8–9), 1053–1062.

770 Baumann, L., Holbech, H., Keiter, S., Kinnberg, K. L., Knörr, S., Nagel, T., & Braunbeck, T.
771 (2013). The maturity index as a tool to facilitate the interpretation of changes in vitellogenin
772 production and sex ratio in the Fish Sexual Development Test. *Aquatic Toxicology*, 128–
773 *129*, 34–42. <https://doi.org/10.1016/j.aquatox.2012.11.016>

774 Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., ... Ortí, G. (2017).
775 Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17(1), 162.
776 <https://doi.org/10.1186/s12862-017-0958-3>

777 Bhanu, S. V., & Philip, B. (2011). Effect of ethanol on branchial adenosine triphosphatases in
778 *Oreochromis mossambicus* (Peters). *Toxicology International*, 18(1), 27.
779 <https://doi.org/10.4103/0971-6580.75849>

780 Borella, M. I., Venturieri, A. R., & Mancera, A. J. M. (2009). Immunocytochemical
781 identification of adenohipophyseal cells in the pirarucu (*Arapaima gigas*), an Amazonian
782 basal teleost. *Fish Physiology and Biochemistry*, 35, 3–16. [https://doi.org/10.1007/s10695-](https://doi.org/10.1007/s10695-008-9254-x)
783 [008-9254-x](https://doi.org/10.1007/s10695-008-9254-x)

784 Bousfield, G. R., & Dias, J. A. (2011). Synthesis and secretion of gonadotropins including

785 structure-function correlates. *Reviews in Endocrine and Metabolic Disorders*, (12), 289–
786 302. <https://doi.org/10.1007/s11154-011-9191-3>

787 Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal RNA-Seq
788 quantification. *Nature Biotechnology*, 34(5), 525–527.

789 Brooks, C. L. (2012). Molecular Mechanisms of Prolactin and Its Receptor. *Endocrine Reviews*,
790 33(4), 504–525. <https://doi.org/10.1210/er.2011-1040>

791 Bry, C., Billard, R., & De Montalembert, G. (1978). Induction de la maturation ovocytaire et de
792 l'ovulation par traitement hormonal chez le brochet (*Esox lucius*). *Bulletin Français de*
793 *Pisciculture*, 271, 21–32. <https://doi.org/10.1051/kmae:1978003>

794 Bryant, D. M., Johnson, K., Di Tommaso, T., Tickle, T., Couger, B. M., Payzin-Dogru, D., ...
795 Whited, J. L. (2017). A tissue-mapped axolotl de novo transcriptome enables identification
796 of limb regeneration factors. *Cell Reports*, 18(3), 762–776.
797 <https://doi.org/10.1016/j.celrep.2016.12.063.A>

798 Bu, G., Liang, X., Li, J., & Wang, Y. (2015). General and Comparative Endocrinology Extra-
799 pituitary prolactin (PRL) and prolactin-like protein (PRL-L) in chickens and zebrafish.
800 *General and Comparative Endocrinology*, 220, 143–153.
801 <https://doi.org/10.1016/j.ygcen.2015.02.001>

802 Bulukin, E., Meucci, V., Minunni, M., Pretti, C., Intorre, L., Soldani, G., & Mascini, M. (2007).
803 An optical immunosensor for rapid vitellogenin detection in plasma from carp (*Cyprinus*
804 *carpio*). *Talanta*, 72(2), 785–790. <https://doi.org/10.1016/j.talanta.2006.12.007>

805 Carreiro, C. R. P., Furtado-Neto, M. A. D. A., Mesquita, P. E. C., & Bezerra, T. A. (2011). Sex
806 determination in the Giant fish of Amazon Basin, *Arapaima gigas* (Osteoglossiformes,
807 *Arapaimatidae*), using laparoscopy. *Acta Amazonica*, 41(3), 415–419.
808 <https://doi.org/10.1590/S0044-59672011000300012>

809 Chatakondi, N. G., & Kelly, A. M. (2013). Oocyte Diameter and Plasma Vitellogenin as
810 Predictive Factors to Identify Potential Channel Catfish, *Ictalurus punctatus*, Suitable for
811 Induced Spawning. *Journal of the World Aquaculture Society*, 44(1), 115–123.
812 <https://doi.org/10.1111/jwas.12001>

813 Chen, Q., Takahashi, Y., Oka, K., & Ma, J. (2016). Functional Differences of Very-Low-Density
814 Lipoprotein Receptor Splice Variants in Regulating Wnt Signaling. *Molecular and Cellular*
815 *Biology*, 36(20), 2645–2654. <https://doi.org/10.1128/MCB.00235-16.Address>

816 Choi, C. Y., & Habibi, H. R. (2003). Molecular cloning of estrogen receptor alpha and
817 expression pattern of estrogen receptor subtypes in male and female goldfish. *Molecular*
818 *and Cellular Endocrinology*, 204(1–2), 169–177. <https://doi.org/S030372070200182X> [pii]

819 Chu-Koo, F., Dugué, R., Alván Aguilar, M., Casanova Daza, A., Alcántara Bocanegra, F.,
820 Chávez Veintemilla, C., ... Nuñez, J. (2009). Gender determination in the Paiche or
821 Pirarucu (*Arapaima gigas*) using plasma vitellogenin, 17 β -estradiol, and 11-ketotestosterone
822 levels. *Fish Physiology and Biochemistry*, 35(1), 125–136. [https://doi.org/10.1007/s10695-](https://doi.org/10.1007/s10695-008-9211-8)
823 008-9211-8

824 Couse, J. F., Lindzey, J., Grandien, K. A. J., Gustafsson, J.-åke, & Korach, K. S. (1997). Tissue
825 Distribution and Quantitative Analysis of Estrogen Receptor- α (ER α) and Estrogen
826 Receptor- β (ER β) Messenger Ribonucleic Acid in the Wild-Type and ER α -Knockout
827 Mouse. *Endocrinology*, 138(11), 4613–4621.

828 Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an
829 overview of genetic, physiological, and environmental influences. *Aquaculture*, 208(3–4),
830 191–364. [https://doi.org/10.1016/S0044-8486\(02\)00057-1](https://doi.org/10.1016/S0044-8486(02)00057-1)

831 Dominguez, G. A., Bisesi, J. H., Kroll, K. J., Denslow, N. D., & Sabo-Attwood, T. (2014).
832 Control of Transcriptional Repression of the Vitellogenin Receptor Gene in Largemouth
833 Bass (*Micropterus Salmoides*) by Select Estrogen Receptors Isotypes. *Toxicological*
834 *Sciences*, 141(2), 423–431. <https://doi.org/10.1093/toxsci/kfu145>

835 Dominguez, G. A., Quattro, J. M., Denslow, N. D., Kroll, K. J., Prucha, M. S., Porak, W. F., ...
836 Sabo-Attwood, T. L. (2012). Identification and Transcriptional Modulation of the
837 Largemouth Bass, *Micropterus salmoides*, Vitellogenin Receptor During Oocyte
838 Development by Insulin and Sex Steroids¹. *Biology of Reproduction*, 87(3), 1–12.
839 <https://doi.org/10.1095/biolreprod.112.099812>

840 Dugué, R., Chu Koo, F., Alcántara Bocanegra, F., Duponchelle, F., Renno, J. F., & Nuñez, J.
841 (2008). Purification and assay of *Arapaima gigas* vitellogenin : Potential use for sex
842 determination by. *Cybium*, 32(2), 111.

843 Ezekiel, E. N., & Abowei, J. F. N. (2013). Length-Weight Relationship and Condition Factor of
844 *Heterotis niloticus* from Amassoma flood plain , Niger Delta , Nigeria . *Applied Science*
845 *Report*, 4(1), 164–172.

846 Faria, M. T., Carvalho, R. F., Sevilhano, T. C. A., Oliveira, N. A. J., Silva, C. F. P., Oliveira, J.
847 E., ... Bartolini, P. (2013). Isolation of the pituitary gonadotrophic α -subunit hormone of
848 the giant amazonian fish: Pirarucu (*Arapaima gigas*). *Fish Physiology and Biochemistry*,
849 39(3), 683–693. <https://doi.org/10.1007/s10695-012-9730-1>

850 Feng, W., Matzuk, M. M., Mountjoy, K., Bedows, E., Ruddon, R. W., & Boime, I. (1995). Ethe
851 Asparagine-linked Oligosaccharides of the Human Chorionic Gonadotropin β Subunit
852 Facilitate Correct Disulfide Bond Pairing. *Journal of Biological Chemistry*, 270(20),

853 11851–11859.

854 Freeman, M. E., Kanyicska, B., Lerant, A., & Nagy, G. (2000). Prolactin: Structure, function,
855 and regulation of secretion. *Physiological Reviews*, *80*(4), 1523–1631.
856 <https://doi.org/10.1016/B978-012515400-0/50037-3>

857 Gao, X. M., Zhou, Y., Zhang, D. D., Hou, C. C., & Zhu, J. Q. (2019). Multiple vitellogenin
858 genes (vtgs) in large yellow croaker (*Larimichthys crocea*): molecular characterization and
859 expression pattern analysis during ovarian development. *Fish Physiology and Biochemistry*,
860 *45*, 829–848. <https://doi.org/10.1007/s10695-018-0569-y>

861 Genovese, G., Regueira, M., Da Cuña, R. H., Ferreira, M. F., Varela, M. L., & Lo Nostro, F. L.
862 (2014). Nonmonotonic response of vitellogenin and estrogen receptor α gene expression
863 after octylphenol exposure of *Cichlasoma dimerus* (Perciformes, Cichlidae). *Aquatic*
864 *Toxicology*, *156*, 30–40. <https://doi.org/10.1016/j.aquatox.2014.07.019>

865 Giguère, V. (1999). Orphan Nuclear Receptors : From Gene to Function *. *Endocrine Reviews*,
866 *20*(5), 689–725.

867 Goffin, V., Shiverick, K. T., Kelly, P. A., & Martial, J. A. (1996). Sequence-Function
868 Relationships Within the Expanding. *Endocrine Reviews*, *17*(4), 385–410.

869 Gordon, M. R., Owen, T. G., Ternan, T. A., & Hildebrand, L. D. (1984). Measurement of a sex-
870 specific protein in skin mucus of premature coho salmon (*Oncorhynchus kisutch*).
871 *Aquaculture*, *43*(1–3), 333–339. [https://doi.org/10.1016/0044-8486\(84\)90034-6](https://doi.org/10.1016/0044-8486(84)90034-6)

872 Grabherr, M., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A.
873 (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq
874 data. *Nature Biotechnology*, *29*(7), 644–652. <https://doi.org/10.1038/nbt.1883>. Trinity

875 Guo-Qing, L., & Wilson, M. V. H. (1996). The discovery of Heterotidinae (Teleostei:
876 Osteoglossidae) from the Paleocene Paskapoo Formation of Alberta, Canada. *Journal of*
877 *Vertebrate Paleontology*, *16*(2), 198–209. <https://doi.org/10.1080/02724634.1996.10011308>

878 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Regev,
879 A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation
880 and analysis with Trinity. In *Nature Protocols* (Vol. 8).
881 <https://doi.org/10.1038/nprot.2013.084>. De

882 Hallerman, E. M. (2006). Use of Molecular Tools for Research and Improvement of Aquaculture
883 Stocks. *Aquaculture*, *58*(4), 286–296.

884 Hara, A., Hiramatsu, N., & Fujita, T. (2016). Vitellogenesis and choriogenesis in fishes.
885 *Fisheries Science*, *82*(2), 187–202. <https://doi.org/10.1007/s12562-015-0957-5>

886 Hawkins, M. B., & Thomas, P. (2004). The unusual binding properties of the third distinct

887 teleost estrogen receptor subtype ER β are accompanied by highly conserved amino acid
888 changes in the ligand binding domain. *Endocrinology*, 145(6), 2968–2977.
889 <https://doi.org/10.1210/en.2003-0806>

890 Hawkins, M. B., Thornton, J. W., Crews, D., Skipper, J. K., Dotte, A., & Thomas, P. (2000).
891 Identification of a third distinct estrogen receptor and reclassification of estrogen receptors
892 in teleosts. *Proceedings of the National Academy of Sciences*, 97(20), 10751–10756.
893 <https://doi.org/10.1073/pnas.97.20.10751>

894 Hilton, E. J. (2001). Tongue Bite Apparatus of Osteoglossomorph Fishes: Variation of a
895 Character Complex. *Copeia*, 2001(2), 372–381.

896 Hilton, E. J. (2003). Comparative osteology and phylogenetic systematics of fossil and living
897 bony-tongue fishes (Actinopterygii, Teleostei, Osteoglossomorpha). *Zoological Journal of*
898 *the Linnean Society*, 137(1), 1–100. <https://doi.org/10.1046/j.1096-3642.2003.00032.x>

899 Hiramatsu, N., Fukadal, H., Sullivan, C. V., & Hara, A. (2001). Simple and Sensitive Detection
900 of Vitellogenin Receptor (s) in Sakhalin Taimen (*Hucho perryi*). *Bull. Fish Sci. Hokkaido*
901 *Univ.*, 52(1), 5–9.

902 Hiramatsu, N., Luo, W., Reading, B. J., Sullivan, C. V., Mizuta, H., Ryu, Y.-W., ... Hara, A.
903 (2013). Multiple ovarian lipoprotein receptors in teleosts. *Fish Physiol Biochem*, 39, 29–32.
904 <https://doi.org/10.1007/s10695-012-9612-6>

905 Hiramatsu, N., Matsubara, T., Fujita, T., Sullivan, C. V., & Hara, A. (2006). Multiple piscine
906 vitellogenins: Biomarkers of fish exposure to estrogenic endocrine disruptors in aquatic
907 environments. *Marine Biology*, 149(1), 35–47. <https://doi.org/10.1007/s00227-005-0214-z>

908 Huang, X., Hui, M. N. Y., Liu, Y., Yuen, D. S. H., Zhang, Y., Chan, W. Y., ... Cheng, C. H. K.
909 (2009). Discovery of a Novel Prolactin in Non-Mammalian Vertebrates : Evolutionary
910 Perspectives and Its Involvement in Teleost Retina Development. *PLoS ONE*, 4(7), e6163.
911 <https://doi.org/10.1371/journal.pone.0006163>

912 IRD. (2008). Vers la domestication du plus grand poisson de l'Amazonie. *Actualité Scientifiques*,
913 306, 1–2.

914 Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., Uchiyama, S., Sasaki, H., ... Chambon, P.
915 (1995). Activation of the Estrogen Receptor Through γ -phosphorylation by Mitogen-
916 Activated Protein Kinase. *Science*, 270, 1491–1494.
917 <https://doi.org/10.16377/j.cnki.issn1007-7731.2011.17.025>

918 Kong, H. J., Kim, J. L., Moon, J. Y., Kim, W. J., Kim, H. S., Park, J. Y., ... An, C. M. (2014).
919 Characterization, expression profile, and promoter analysis of the *Rhodeus uyekii*
920 vitellogenin A α 1 gene. *International Journal of Molecular Sciences*, 15(10), 18804–18818.

921 <https://doi.org/10.3390/ijms151018804>

922 Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA : Fast and accurate filtering of
923 BIOINFORMATICS SortMeRNA : Fast and accurate filtering of ribosomal RNAs in
924 metatranscriptomic data. *Bioinformatics*, 28(24), 3211–3217.
925 <https://doi.org/10.1093/bioinformatics/bts611>

926 Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting
927 transmembrane protein topology with a hidden Markov model: Application to complete
928 genomes. *Journal of Molecular Biology*, 305(3), 567–580.
929 <https://doi.org/10.1006/jmbi.2000.4315>

930 Kumar, R., Zakharov, M. N., Khan, S. H., Miki, R., Jang, H., Toraldo, G., ... Jasuja, R. (2011).
931 The Dynamic Structure of the Estrogen Receptor. *Journal of Amino Acids*, 2011, 1–7.
932 <https://doi.org/10.4061/2011/812540>

933 Kumazawa, Y., & Nishida, M. (2000). Molecular Phylogeny of Osteoglossoids : A New Model
934 for Gondwanian Origin and Plate Tectonic Transportation of the Asian Arowana. *Molecular*
935 *Biology and Evolution*, 17(12), 1869–1878.

936 Lavoué, S. (2016). Was Gondwanan breakup the cause of the intercontinental distribution of
937 Osteoglossiformes? A time-calibrated phylogenetic test combining molecular,
938 morphological, and paleontological evidence. *Molecular Phylogenetics and Evolution*, 99,
939 34–43. <https://doi.org/10.1016/j.ympev.2016.03.008>

940 Lavoué, S., & Sullivan, J. P. (2004). Simultaneous analysis of five molecular markers provides a
941 well-supported phylogenetic hypothesis for the living bony-tongue fishes
942 (Osteoglossomorpha: Teleostei). *Molecular Phylogenetics and Evolution*, 33(1), 171–185.
943 <https://doi.org/10.1016/j.ympev.2004.04.021>

944 Lazard, J. (1990). Transferts de poissons et développement de la production piscicole. *Revue*
945 *Hydrobiologie Tropicale*, 23(3), 251–265.

946 Le Menn, F. (1979). Induction of vitellogenin by estradiol and androgens in a teleostean fish:
947 *Gobius niger* L. *Comptes Rendus Des Seances de l'Academie Des Sciences*, 289(4), 413–
948 416. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/117947>

949 Leung, Y.-K., Mak, P., Hassan, S., & Ho, S. (2006). Estrogen receptor (ER) beta isoforms : A
950 key to understanding ER beta signaling. *PNAS*, 103(35), 13162–13167.

951 Li, A., Sadasivam, M., & Ding, J. L. (2003). Receptor-ligand interaction between vitellogenin
952 receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and
953 apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-
954 terminal region of Vtg. *Journal of Biological Chemistry*, 278(5), 2799–2806.

955 <https://doi.org/10.1074/jbc.M205067200>

956 Luo, W., Ito, Y., Mizuta, H., Massaki, K., Hiramatsu, N., Todo, T., ... Hara, A. (2013).
957 Molecular cloning and partial characterization of an ovarian receptor with seven ligand
958 binding repeats, an orthologue of low-density lipoprotein receptor, in the cutthroat trout
959 (*Oncorhynchus clarki*). *Comparative Biochemistry and Physiology - A Molecular and*
960 *Integrative Physiology*, 166(2), 263–271. <https://doi.org/10.1016/j.cbpa.2013.06.026>

961 Mañanós, E., Núñez-Rodríguez, J., Le Menn, F., Zanuy, S., & Carrillo, M. (2007). Identification
962 of vitellogenin receptors in the ovary of a teleost fish, the Mediterranean sea bass
963 (*Dicentrarchus labrax*). *Reproduction Nutrition Development*, 37, 51–61.
964 <https://doi.org/10.1051/rnd:19970106>

965 Manzon, L. A. (2002). The Role of Prolactin in Fish Osmoregulation : A Review. *General and*
966 *Comparative Endocrinology*, 125, 291–310. <https://doi.org/10.1006/gcen.2001.7746>

967 Marcos, P. L., & Adalberto, L. V. (2015). Differentially expressed genes in the pituitary of the
968 Amazonian fish *Arapaima gigas*. *International Journal of Fisheries and Aquaculture*, 7(8),
969 132–141. <https://doi.org/10.5897/IJFA15.0473>

970 Marlatt, V. L., Martyniuk, C. J., Zhang, D., Xiong, H., Watt, J., Xia, X., ... Trudeau, V. L.
971 (2008). Auto-regulation of estrogen receptor subtypes and gene expression profiling of 17 β -
972 estradiol action in the neuroendocrine axis of male goldfish. *Molecular and Cellular*
973 *Endocrinology*, 283(1–2), 38–48. <https://doi.org/10.1016/j.mce.2007.10.013>

974 Maruska, K. P., & Fernald, R. D. (2011). Social Regulation of Gene Expression in the
975 Hypothalamic-Pituitary-Gonadal Axis. *Physiology*, 26(6), 412–423.
976 <https://doi.org/10.1152/physiol.00032.2011>

977 Mehdi, Y., & Ehsan, S. (2013). A review of the control of reproduction and hormonal
978 manipulations in finfish species. *International Journal of Agricultural Research*, 1(1), 15–
979 21.

980 Meier, F., Beck, S., Grassl, N., Lubeck, M., Park, M. A., Raether, O., & Mann, M. (2015).
981 Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and
982 Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *Journal of Proteome*
983 *Research*, 14(12), 5378–5387. <https://doi.org/10.1021/acs.jproteome.5b00932>

984 Métivier, R., Stark, A., Flouriot, G., Hübner, M. R., Brand, H., Penot, G., ... Gannon, F. (2002).
985 A Dynamic Structural Model for Estrogen Receptor alpha Activation by Ligands ,
986 Emphasizing the Role of Interactions between Distant A and E Domains. *Molecular Cell*,
987 10, 1019–1032.

988 Meucci, V., & Arukwe, A. (2005). Detection of vitellogenin and zona radiata protein expressions

989 in surface mucus of immature juvenile Atlantic salmon (*Salmo salar*) exposed to waterborne
990 nonylphenol. *Aquatic Toxicology (Amsterdam, Netherlands)*, 73(1), 1–10.
991 <https://doi.org/10.1016/j.aquatox.2005.03.021>

992 Mizuta, H., Luo, W., Ito, Y., Mushiobira, Y., Todo, T., Hara, A., ... Hiramatsu, N. (2013).
993 Ovarian expression and localization of a vitellogenin receptor with eight ligand binding
994 repeats in the cutthroat trout (*Oncorhynchus clarki*). *Comparative Biochemistry and*
995 *Physiology. Part B, Biochemistry & Molecular Biology*, 166(1), 81–90.
996 <https://doi.org/10.1016/j.cbpb.2013.07.005>

997 Moncaut, N., Nostro, F. Lo, & Maggese, M. C. (2003). Vitellogenin detection in surface mucus
998 of the South American cichlid fish *Cichlasoma dimerus* (Heckel, 1840) induced by
999 estradiol-17beta. Effects on liver and gonads. *Aquatic Toxicology (Amsterdam,*
1000 *Netherlands)*, 63(2), 127–137. [https://doi.org/10.1016/S0166-445X\(02\)00175-3](https://doi.org/10.1016/S0166-445X(02)00175-3)

1001 Monentcham, S. E., Kouam, J., Pouomogne, V., & Kestemont, P. (2009). Biology and prospect
1002 for aquaculture of African bonytongue, *Heterotis niloticus* (Cuvier, 1829): A review.
1003 *Aquaculture*, 289(3–4), 191–198. <https://doi.org/10.1016/j.aquaculture.2009.01.019>

1004 Moreau, J. (1982). Exposé synoptique des données biologiques sur *Heterotis niloticus* (Cuvier,
1005 1829). *FAO Synop. Pêches*, 131, 1–45.

1006 Mushiobira, Y., Mizuta, H., Luo, W., Todo, T., Hara, A., Reading, B. J., ... Hiramatsu, N.
1007 (2015). Molecular cloning and partial characterization of a low-density lipoprotein receptor-
1008 related protein 13 (Lrp13) involved in vitellogenin uptake in the cutthroat trout
1009 (*Oncorhynchus clarki*). *Molecular Reproduction and Development*, 82(12), 986–1000.
1010 <https://doi.org/10.1002/mrd.22579>

1011 Mylonas, C. C., Fostier, A., & Zanuy, S. (2010). Reproduction and Broodstock Management
1012 General and Comparative Endocrinology Broodstock management and hormonal
1013 manipulations of fish reproduction. *General and Comparative Endocrinology*, 165, 516–
1014 534. <https://doi.org/10.1002/9781444392210.ch4>

1015 Ndiaye, P., Forgue, J., Lamothe, V., Cauty, C., Tacon, P., Lafon, P., ... Núñez, J. (2006). Tilapia
1016 (*Oreochromis niloticus*) vitellogenins: development of homologous and heterologous
1017 ELISAs and analysis of vitellogenin pathway through the ovarian follicle. *Journal of*
1018 *Experimental Zoology Part A: Comparative Experimental Biology*, 305A(7), 576–593.
1019 <https://doi.org/10.1002/jez.a.290>

1020 Nelson, E. R., & Habibi, H. R. (2010). Functional significance of nuclear estrogen receptor
1021 subtypes in the liver of goldfish. *Endocrinology*, 151(4), 1668–1676.
1022 <https://doi.org/10.1210/en.2009-1447>

- 1023 Nelson, E. R., & Habibi, H. R. (2013). Estrogen receptor function and regulation in fish and
1024 other vertebrates. *General and Comparative Endocrinology*, *192*, 15–24.
1025 <https://doi.org/10.1016/j.ygcen.2013.03.032>
- 1026 Nelson, E. R., Wiehler, W. B., Cole, W. C., & Habibi, H. R. (2007). Homologous Regulation of
1027 Estrogen Receptor Subtypes in Goldfish (*Carassius auratus*). *Molecular Reproduction and*
1028 *Development*, *74*(1), 1105–1112. <https://doi.org/10.1002/mrd>
- 1029 Nelson, G. J. (1968). Gill arches of teleostean fishes of the division Osteoglossomorpha. *Journal*
1030 *of the Linnean Society of London, Zoology*, *47*(312), 261–277.
1031 <https://doi.org/10.1111/j.1096-3642.1968.tb00511.x>
- 1032 Nelson, G. J. (1969). Infraorbital bones and their bearing on the phylogeny and geography of
1033 Osteoglossomorph fishc. *American Museum Novitates*, *2394*, 1–37.
- 1034 Norris, D. O., & Lopez, K. H. (2011). Hormones and reproduction of vertebrates, Volume 1 -
1035 Fishes. In *Saudi Med J* (Vol. 1).
- 1036 Noso, T., Nicoll, C. S., Polenov, A. L., & Kawauchi, H. (1993). The primary structure of
1037 sturgeon prolactin: Phylogenetic implication. *General and Comparative Endocrinology*,
1038 Vol. 91, pp. 90–95. <https://doi.org/10.1006/gcen.1993.1108>
- 1039 Núñez-Rodríguez, J., Bon, E., & Le Menn, F. (1996). Vitellogenin receptors during
1040 Vitellogenesis in the rainbow trout *Oncorhynchus mykiss*. *The Journal of Experimental*
1041 *Zoology*, *274*, 163–170. [https://doi.org/10.1002/\(sici\)1097-010x\(19960215\)274:3<163::aid-](https://doi.org/10.1002/(sici)1097-010x(19960215)274:3<163::aid-)
1042 [jez3>3.0.co;2-m](https://doi.org/10.1002/(sici)1097-010x(19960215)274:3<163::aid-jez3>3.0.co;2-m)
- 1043 Nuñez-Rodriguez, J., Kah, O., Geffard, M., & Le Menn, F. (1989). Enzyme Linked
1044 Immunosorbent Assay (ELISA) for sole (*Solea vulgaris*) vitellogenin. *Comparative*
1045 *Biochemistry and Physiology*, *92B*(4), 741–746.
- 1046 Nuñez, J. (2008). Arapaima gigas se dévoile. *Science Au Sud*, *47*, 1.
- 1047 Nunez Rodriguez, J., Dugue, R., Otome, Z. J., Hem, S., & Le Menn, F. (1997). Vitellogenin
1048 plasma levels in two cultured African catfish species, *Chrysichthys nigrodigitatus*
1049 (*Claroteidae*) and *Heterobranchus longifilis* (*Clariidae*). *Aquat. Living Resources*, *10*, 231–
1050 238.
- 1051 Odo, G., Nwamba, H., & Eyo, J. (2009). Aspects Of The Biology Of *Heterotis niloticus* Cuvier
1052 1829 (Osteoglossiformes: Osteoglossidae) In The Anambra Flood River System, Nigeria.
1053 *Animal Research International*, *6*(2), 994–1002. <https://doi.org/10.4314/ari.v6i2.48131>
- 1054 Oladosu, G. A., Obi, A., & Oladosu, O. O. (2007). Sex determination in *Heterotis niloticus*
1055 (Cuvier 1829) Based on morphometric features. *ASSET Series B*, *6*(1), 22–30.
- 1056 Om, A. D., Jasmani, S., Ismail, N., Yeong, S. Y., & Abol-Munafi, A. B. (2013). Application

1057 MALDI TOF on protein identification of vitellogenin in giant grouper (*Epinephelus*
1058 *lanceolatus*). *Fish Physiology and Biochemistry*, 39(5), 1277–1286.
1059 <https://doi.org/10.1007/s10695-013-9782-x>

1060 Pelissero, C., Cuisset, B., & Le Menn, F. (1989). The influence of sex steroids in commercial
1061 meals and fish diets on plasma concentration of estrogens and vitellogenin in cultured
1062 Siberian sturgeon *Acipense baeri*. *Aquatic Living Resources*, 2, 161–168.
1063 <https://doi.org/10.1051/alr>

1064 Pelissero, C., & Sumpter, J. P. (1992). Steroids and “steroid-like” substances in fish diets.
1065 *Aquaculture*, 107, 283–301. [https://doi.org/10.1016/0044-8486\(92\)90078-Y](https://doi.org/10.1016/0044-8486(92)90078-Y)

1066 Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating
1067 signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785–786.
1068 <https://doi.org/10.1038/nmeth.1701>

1069 Plant, T. M. (2015). The hypothalamo-pituitary-gonadal axis. *Journal of Endocrinology*, 226(2),
1070 T41–T54. <https://doi.org/10.1530/JOE-15-0113>

1071 Pousis, C., Santamaria, N., Zupa, R., De Giorgi, C., Mylonas, C. C., Bridges, C. R., ... Corriero,
1072 A. (2012). Expression of vitellogenin receptor gene in the ovary of wild and captive
1073 Atlantic bluefin tuna (*Thunnus thynnus*). *Animal Reproduction Science*, 132, 101–110.
1074 <https://doi.org/10.1016/j.anireprosci.2012.03.014>

1075 Prat, F., Coward, K., Sumpter, J. P., & Tyler, C. R. (1998). Molecular Characterization and
1076 Expression of two Ovarian Lipoprotein Receptors in the Rainbow Trout, *Oncorhynchus*
1077 *mykiss* 1. *Biology of Reproduction*, 58(5), 1146–1153.
1078 <https://doi.org/10.1095/biolreprod58.5.1146>

1079 Reading, B. J., Hiramatsu, N., Schilling, J., Molloy, K. T., Glassbrook, N., Mizuta, H., ...
1080 Sullivan, C. V. (2014). Lrp13 is a novel vertebrate lipoprotein receptor that binds
1081 vitellogenins in teleost fishes. *Journal of Lipid Research*, 55(11), 2287–2295.
1082 <https://doi.org/10.1194/jlr.M050286>

1083 Reading, B. J., Hiramatsu, N., & Sullivan, C. V. (2011). Disparate Binding of Three Types of
1084 Vitellogenin to Multiple Forms of Vitellogenin Receptor in White Perch 1. *Biology of*
1085 *Reproduction*, 84, 392–399. <https://doi.org/10.1095/biolreprod.110.087981>

1086 Roy, R. L., Morin, Y., Courtenay, S. C., & Robichaud, P. (2004). Purification of vitellogenin
1087 from smooth flounder (*Pleuronectes putnami*) and measurement in plasma by homologous
1088 ELISA. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular*
1089 *Biology*, 139(2), 235–244. <https://doi.org/10.1016/j.cbpc.2004.07.006>

1090 Sabo-Attwood, T., Kroll, K. J., & Denslow, N. D. (2004). Differential expression of largemouth

1091 bass (*Micropterus salmoides*) estrogen receptor isotypes alpha, beta, and gamma by
1092 estradiol. *Molecular and Cellular Endocrinology*, 218(1–2), 107–118.
1093 <https://doi.org/10.1016/j.mce.2003.12.007>

1094 Sappington, T. W., & Raikhel, A. S. (1998). Ligand-binding domains in vitellogenin receptors
1095 and other LDL-receptor family members share a common ancestral ordering of cysteine-
1096 rich repeats. *Journal of Molecular Evolution*, 46(4), 476–487.
1097 <https://doi.org/10.1007/PL00006328>

1098 Schilling, J., Loziuk, P. L., Muddiman, D. C., Daniels, H. V., & Reading, B. J. (2015).
1099 Mechanisms of egg yolk formation and implications on early life history of white perch
1100 (*Morone americana*). *PLoS ONE*, 10(11), 1–20.
1101 <https://doi.org/10.1371/journal.pone.0143225>

1102 Sevilhano, T., Carvalho, R. F. De, Oliveira, N. A. D. J., Oliveira, J. E., Maltarollo, V. G.,
1103 Trossini, G., ... Bartolini, P. (2017). Molecular cloning and characterization of pirarucu (
1104 *Arapaima gigas*) follicle-stimulating hormone and luteinizing hormone β -subunit cDNAs.
1105 *PLoS ONE*, 12(8), e0183545.

1106 Skorupski, J., & Kmiec, M. (2013). Structural and functional characteristics of prolactin and its
1107 gene in the American mink (*Neovison vison* Schreb., 1777). A review / Charakterystyka
1108 strukturalna i funkcjonalna prolaktyny i jej genu u norki amerykańskiej (*Neovison vison*
1109 schreb., 1777). *Pra. Annales UMCS, Zootechnica*, 30(4). [https://doi.org/10.2478/v10083-](https://doi.org/10.2478/v10083-012-0035-8)
1110 [012-0035-8](https://doi.org/10.2478/v10083-012-0035-8)

1111 Šlechtová, T., Gilar, M., Kalíková, K., & Tesařová, E. (2015). Insight into Trypsin Miscleavage:
1112 Comparison of Kinetic Constants of Problematic Peptide Sequences. *Analytical Chemistry*,
1113 87(15), 7636–7643. <https://doi.org/10.1021/acs.analchem.5b00866>

1114 So, W., Kwok, H., & Ge, W. (2005). Zebrafish Gonadotropins and Their Receptors : II . Cloning
1115 and Characterization of Zebrafish Follicle-Stimulating Hormone and Luteinizing Hormone
1116 Subunits — Their Spatial-Temporal Expression Patterns and Receptor Specificity. *Biology*
1117 *of Reproduction*, 72, 1382–1396. <https://doi.org/10.1095/biolreprod.104.038216>

1118 Stifani, S., Le Menn, F., Nunez Rodriguez, J., & Schneider, W. J. (1990). Regulation of
1119 oogenesis: The piscine receptor for vitellogenin. *Biochimica et Biophysica Acta (BBA) -*
1120 *Lipids and Lipid Metabolism*, 1045, 271–279. [https://doi.org/10.1016/0005-2760\(90\)90130-](https://doi.org/10.1016/0005-2760(90)90130-P)
1121 [P](https://doi.org/10.1016/0005-2760(90)90130-P)

1122 Takahashi, A., Kanda, S., Abe, T., & Oka, Y. (2016). Evolution of the hypothalamic-pituitary-
1123 gonadal axis regulation in vertebrates revealed by knockout medaka. *Endocrinology*,
1124 157(10), 3994–4002. <https://doi.org/10.1210/en.2016-1356>

- 1125 Taranger, G. L., Carrillo, M., Schulz, R. W., Fontaine, P., Zanuy, S., Felip, A., ... Hansen, T.
1126 (2010). Control of puberty in farmed fish. *General and Comparative Endocrinology*,
1127 165(3), 483–515. <https://doi.org/10.1016/j.ygcen.2009.05.004>
- 1128 Teilum, K., Hoch, J. C., Goffin, V., Kinet, S., Martial, J. A., & Kragelund, B. B. (2005). Solution
1129 Structure of Human Prolactin. *Journal of Molecular Biology*, 351, 810–823.
1130 <https://doi.org/10.1016/j.jmb.2005.06.042>
- 1131 Trichet, V., Buisine, N., Mouchel, N., Morán, P., Pendás, A. M., Le Pennec, J. P., & Wolff, J.
1132 (2000). Genomic analysis of the vitellogenin locus in rainbow trout (*Oncorhynchus mykiss*)
1133 reveals a complex history of gene amplification and retroposon activity. *Molecular and*
1134 *General Genetics*, 263(5), 828–837. <https://doi.org/10.1007/s004380000247>
- 1135 Trommsdorff, M., Gotthardt, M., Hiesberger, T., Shelton, J., Stockinger, W., Nimpf, J., ... Herz,
1136 J. (1999). Reeler/disabled-like disruption of neuronal migration in knockout mice lacking
1137 the VLDL receptor and ApoE receptor 2. *Cell*, 97(6), 689–701.
1138 [https://doi.org/10.1016/S0092-8674\(00\)80782-5](https://doi.org/10.1016/S0092-8674(00)80782-5)
- 1139 Uchida, K., Moriyama, S., Chiba, H., Shimotani, T., Honda, K., Miki, M., ... Nozaki, M. (2010).
1140 Evolutionary origin of a functional gonadotropin in the pituitary of the most primitive
1141 vertebrate, hagfish. *PNAS*, 107(36), 15832–15837.
1142 <https://doi.org/10.1073/pnas.1002208107>
- 1143 Ulloa-Aguirre, A., Timossi, C., Damián-matsumura, P., & Dias, J. A. (1999). Role of
1144 Glycosylation in Function of Follicle-Stimulating Hormone. *Endocrine*, 11, 205–215.
1145 <https://doi.org/10.1385/ENDO>
- 1146 Unal, G., Marquez, E. C., Feld, M., Stavropoulos, P., & Callard, I. P. (2014). Isolation of
1147 estrogen receptor subtypes and vitellogenin genes: Expression in female *Chalcalburnus*
1148 *tarichi*. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular*
1149 *Biology*, 172–173(1), 67–73. <https://doi.org/10.1016/j.cbpb.2014.04.002>
- 1150 Vischer, H. F., Marques, R. B., Granneman, J. C. M., Linskens, M. H. K., Schulz, R. W., &
1151 Bogerd, J. (2004). Receptor-selective determinants in catfish gonadotropin seat-belt loops.
1152 *Molecular and Cellular Endocrinology*, 224, 55–63.
1153 <https://doi.org/10.1016/j.mce.2004.06.011>
- 1154 Wallace, R. A., Hoch, K. L., & Carnevali, O. (1990). Placement of small lipovitellin subunits
1155 within the vitellogenin precursor in *Xenopus laevis*. *Journal of Molecular Biology*, 213(3),
1156 407–409. [https://doi.org/10.1016/S0022-2836\(05\)80203-7](https://doi.org/10.1016/S0022-2836(05)80203-7)
- 1157 Wang, H., Tan, J. T. T., Emelyanov, A., Korzh, V., & Gong, Z. (2005). Hepatic and extrahepatic
1158 expression of vitellogenin genes in the zebrafish, *Danio rerio*. *Gene*, 356(1–2), 91–100.

1159 <https://doi.org/10.1016/j.gene.2005.03.041>

1160 Wang, N., Wang, R., Hu, Q., Xu, W., Zhu, Y., Yan, F., & Chen, S. (2017). Characterization of a
1161 low-density lipoprotein receptor, Lrp13, in Chinese tongue sole (*Cynoglossus semilaevis*)
1162 and medaka (*Oryzias latipes*). *Fish Physiology and Biochemistry*, *43*(5), 1289–1298.
1163 <https://doi.org/10.1007/s10695-017-0372-1>

1164 Wang, R., Gao, Y., Zhang, L., Zhang, Y., Fang, Z., He, J., ... Ma, G. (2010). Cloning,
1165 expression, and induction by 17- β estradiol (E2) of a vitellogenin gene in the white cloud
1166 mountain minnow *Tanichthys albonubes*. *Fish Physiology and Biochemistry*, *36*(2), 157–
1167 164. <https://doi.org/10.1007/s10695-008-9222-5>

1168 Watahiki, M., Yamamoto, M., Yamakawa, M., Tanaka, M., & Nakashima, K. (1989). Conserved
1169 and Unique Amino Acid Residues in the Domains of the Growth Hormones. *Journal of*
1170 *Biological Chemistry*, *264*(1), 312–316.

1171 Whittington, C. M., & Wilson, A. B. (2013). General and Comparative Endocrinology The role
1172 of prolactin in fish reproduction. *General and Comparative Endocrinology*, *191*, 123–136.
1173 <https://doi.org/10.1016/j.ygcen.2013.05.027>

1174 Williams, V. N., Reading, B. J., Hiramatsu, N., Amano, H., Glassbrook, N., Hara, A., &
1175 Sullivan, C. V. (2014). Multiple vitellogenins and product yolk proteins in striped bass,
1176 *Morone saxatilis*: molecular characterization and processing during oocyte growth and
1177 maturation. *Fish Physiology and Biochemistry*, *40*(2), 395–415.
1178 <https://doi.org/10.1007/s10695-013-9852-0>

1179 Yang, H. M., Wang, Y., Wang, Z. Y., & Wang, X. X. (2017). Seasonal and photoperiodic
1180 regulation of reproductive hormones and related genes in Yangzhou geese. *Poultry Science*,
1181 *96*(2), 486–490. <https://doi.org/10.3382/ps/pew340>

1182 Yasuda, A., Kawauchi, H., & Papkoff, H. (1990). The complete amino acid sequence of
1183 prolactin from the sea turtle (*Chelonia mydas*). *General and Comparative Endocrinology*,
1184 *80*(3), 363–371. [https://doi.org/10.1016/0016-6480\(90\)90185-O](https://doi.org/10.1016/0016-6480(90)90185-O)

1185 Yilmaz, O., Prat, F., Ibañez, A. J., Amano, H., Koksoy, S., & Sullivan, C. V. (2015). Estrogen-
1186 induced yolk precursors in European sea bass, *Dicentrarchus labrax*: Status and perspectives
1187 on multiplicity and functioning of vitellogenins. *General and Comparative Endocrinology*,
1188 *221*, 16–22. <https://doi.org/10.1016/j.ygcen.2015.01.018>

1189 Zatylny-Gaudin, C., Cornet, V., Leduc, A., Zanuttini, B., Corre, E., Le Corguillé, G., ... Henry,
1190 J. (2016). Neuropeptidome of the Cephalopod *Sepia officinalis* : Identification, Tissue
1191 Mapping, and Expression Pattern of Neuropeptides and Neurohormones during Egg Laying.
1192 *Journal of Proteome Research*, *15*, 48–67. <https://doi.org/10.1021/acs.jproteome.5b00463>

- 1193 Zechel, C., Shen, X., Chen, J., Chen, Z., Chambon, P., & Gronemeyerl, H. (1994). The
1194 dimerization interfaces formed between the DNA binding domains of RXR , RAR and TR
1195 determine the binding specificity and polarity of the full-length receptors to direct repeats.
1196 *The EMBO Journal*, 13(6), 1425–1433.
- 1197 Zhang, S., Dong, Y., & Cui, P. (2015). Vitellogenin is an immunocompetent molecule for
1198 mother and offspring in fish. *Fish & Shellfish Immunology*, 46(2), 710–715.
1199 <https://doi.org/10.1016/j.fsi.2015.08.011>
- 1200 Zhong, L., Yuan, L., Rao, Y., Li, Z., Zhang, X., Liao, T., ... Dai, H. (2014). Distribution of
1201 vitellogenin in zebrafish (*Danio rerio*) tissues for biomarker analysis. *Aquatic Toxicology*
1202 (*Amsterdam, Netherlands*), 149, 1–7. <https://doi.org/10.1016/j.aquatox.2014.01.022>
- 1203 Zohar, Y., Muñoz-Cueto, J. A., Elizur, A., & Kah, O. (2010). Neuroendocrinology of
1204 reproduction in teleost fish. *General and Comparative Endocrinology*, 165(3), 438–455.
1205 <https://doi.org/10.1016/j.ygcen.2009.04.017>

1 M K A A V F A L A L A L A L V A G ▽ Q Q N I L T P D F A A K K T Y V Y K Y E
 2 **ATG**AAAGCAGCTGTTTTCTGACTGGCTGGCCCTTTGGGGTGGTCAGCAAAAACCTCTAACACCTGACTTTGGCTGTAAGAGACCTATGTGTACAAGTATGAA
 36 A Q L H S E L P E E G L A K A G L K I A S K V L I S R A D Q N T Y L L
 106 GCACAGCTCCATAGTGAACCTACCTGAGGAGGGTCTGGCCAGGCTGGACTGAAGATGCCAGCAAAGTTTCTCATCAGTCCGAGTCCAGGACCACTATTACTC
 71 K L K E P Q I F E Y S G I R P R D D E K P A A K L T Q A L A A Q L L I
 211 AAGCTCAAGAACCTCAGATCTTTGAGTACAGCGGAATACGGCCAGAGACGATTTCAAACCTGCTGCAAAACTCACCCAGGCGCTTGGCCGCTCAGCTCTTATT
 106 P V K F E Y I R G V V G K V F A P A A V S K T I L N L H R G I L N I L
 316 CCAGTCAAGTTGTAATACATCAGGGGTGTGTGGGCAAGGTGTTGCCCCTGCCGAGTCTCCAAAACCTATTCTCAACCTCCACAGAGGAATCTCAACATCTC
 141 H V N I K K T Q N V Y E L Q E A G S Q G V C K T D Y A I S E D T K A E
 421 CATGTCAACATCAAAAAGACAGAAATGTTTACGAGTTGCAGGAGGCTGCATCCAGGGAGTCTGCAAAACTGACTACGCAATCAGTGGAGTACCAAGGCTGAA
 176 H I Y V T K S K D L G N C Q N R I M A D I G I A Y T E T C V Q C Q M R
 526 CACATCTATGTGACCAGTCCAAGATCTGGGGAAGTCCGAGAACAGGATCATGGCAGACATTTGGAATAGTCTACACAGAGACTTGTGTCCAGTCCAGATGAGA
 211 N K N L R G A A T Y S Y V M K P T E T G A L I T K A T Y Q C E V H Q F T
 631 AACAGAATCTGAGAGGACTGCGACTACAGCTACAGTCAATGAGCCACAGCAACAGTGTCTTATCACCAAGGACAGTCCAGGAGGCTTCTCAGTCTCACT
 246 P F H E L R G A G Q V E T R Q E M T F E V E T Q N D P V Q P A H A D A Y V
 736 CTTTCCACGAGCTGAGAGCGGGCAACAGTAAACAAGAGATGACTTTTGTAACCCAGAATGACCCAGTCAACCTGCAACCTGCAACAGGACTCTGAGT
 281 A R G S L Q Y E F A S E L L Q T P L P L M K I T D I K T Q I E E I L N
 841 GCACGTGGATCCCGCAGATGAATTTGCATCTGACTCTCTCAGACACCGCTGCTGATGAAAATCAGTATATCAAGACACAGATGAAGAGATCTTGAAT
 316 H L V L K N N M G E A P L K F L Q L T Q L L R E A K Y E I M N G
 946 CATCTGGTCAAAAACAACATGGGTGAGGTTTACGGAAGATGCTCCGCTAAAATTTCTCAGCTGACTCAACTCTGCGTGAGGCCAAAATATGAGATCATGAATGGC
 351 I W T V T K S K P V F R R W F L D T V P A V G N Q D A V R F I K E K F
 1051 ATTTGAGCTCAAGTCAAATCAAACAGTATTACAGCTGCTGGTTCCTGGATACGGTTCCCTGCCGTGGAATCAAAGACAGTGTGAGGTTTCATCAAGGAGAATTT
 386 I A G D I S A A E T A Q A L L V A L H L L E A N M D T V N L A G T L V
 1126 ATAGCTGGTATATCTCTGCAAGGAGAACCTGCCCGCTACTGGTTGCACTGATCTGCTGAGAAGTAACTGGACCACTCAACTTAGCTGGCAGCTTGGTAT
 421 F H A K A T L R Q S H P M L R E I A M L L G Y G S L V F K L C T K Q G N C P A
 1261 TTCACGCCAAACTCAGTCCACCCTATGTTACGTAAATAGCCATGCTTTGGTTATGGTTTCCTTGGTGTCAAGTTGTGCATAAACACAGAACTGCCCGGCT
 456 E V I K P V H D I A A E A I S K A N V E E I A L V L K V L G N A G H P
 1366 GAAGTCA1AAACCTGTTCCAGATATTGCTGCAGAGCAATTAGTAAAGCGAATGTTGGAAGAGATGCCCCTGGTGCTGAAGGTTTCTGGGAATGACGAGCCACCG
 491 A S I K P I M K V L P G F G S T A S S L T V K V H V D A V M A L R H I
 1471 GCCAGCATCAAGCCGATCATGAAGTCTGCTGCCTGGATTGGGCTACAGCCCTCCTCACTTACAGTGAAGTCCACGTTGATGGGTTGAGACACATT
 526 A K R E N H K V Q E V A L Q L F M N R D L H P E V R M V A C V V L F E
 1576 GCCAAAAGGAAAACCACAAAGTCCAGGAAGTTGCCCTGCAGTGTTCATGAACAGAGATCTCCACCAGAGTGGCAGTGGTGTGTGCTGCTGTTGAG
 561 C K P P I G L V A M I A S A L Q N E K S L Q V A S F A Y S H M K A L T
 1681 TGCAACCTCCATCGGCTGGTAGCTATGATGCAAGTGTCTACAGAAGCAGAAAAGCCTGCAAGTGGCCAGCTTTGCTATTCATATGAAGGCCCTTGACC
 596 R S T A P E L A Q V A A A C N V A I K I L S P K L D R L S Y R F S K T
 1786 AGGAGCACCGCCCAAGTGGCCAAAGTGCAGCTGCTGAATGTTGCCATCAAGATCTTGAAGCCCAAGTTAGATCGACTGAGCTATCGCTTCAGCAAAAAC
 631 I H M D E F F N Y H L M A G A A A T A H L I N D A A S I L P R A L V A K
 1891 ATCCACATGACTTTTTAACTATCAGTTAGTGGCTGCGCTGCAGTCTGCCATTGATCAATGAGCTGCCAGATTTGGCAAGAGCCGTTTGGCAAGGCCGCTTGGGCAAAA
 666 I R A C M L G A D A D V L E F G V R T E G L Q L E A L L K S P A A D P N
 1996 ATCCGTCCTGCATGTGTGGGGCAGCTGCAGATGTGCTGGAGTTGGTGAGAACTGAAGGACTCCAGGAGGCTCTTTGAAAAGTCCAGCTGCGGATCCATAA
 701 A D R L T R M R W R V L N T L K N W K S L P A S Q P V A S Q P V A S A Y L L A A T T
 2101 GCTGACAGGCTCACCCGATGGCGGTGTTCTAAACACTCTGA AAAA CTGGAAGTCACTACCCAGCAAGCAACAGTGGCTTCTGCTTACTTGAAATGGTGTGGC
 736 Q E I A F A N I D R D I E K A V E F A T G A A Q P A L L K N I L N
 2206 CAAGAAATGCTTTGCCAACTCAGACGAGCATCATCGAAAAGCCGTCGAGTTCGCCACAGGAGCTGCTGCACGCCAGCACTGTTGAGAACATCCCTGAAAT
 771 M K P L Q I D I Q I A K P L M T M E V R R I E P T S M G F E I S L
 2311 ATGATCGACTGCTGATTTGATATCCAATTTGCCAAACCCTGATGACCTCTGAGTGGCTGCGCATCTTTCCACGCTTATGGGTTTCCCCTATAGATCAGCCCTC
 806 Y S A A V A A A I V K A K A I L N P K P S D N F R I A Q L M N T D I Q
 2416 TACTCAGCTGCTGGCTCAGCTATGATTAAGCTAAAGCAATCTGAATCCAAAACCTCCGACAACCTCAGGATTTGCTCAGTGTGATGAACACCCGATCTCAG
 841 L D T R V V P S I A V H K Y A V M G V N T A L I Q A A I E A K V K V
 2521 TTGGACCCCGTGTTGACCAAGCATAGCTGTACATAAATATGACGATTTGGGAGTAAACACTGCAGCTCATTGAGCTGCAATTTGAGGCCAAAGTCAAAGTTCAG
 876 K V L P L K F N A R I N I A Q G H Y K I E T V P L F H A Q E R I L D L H
 2626 AAGTCTCTCCCTCAAGTTCAACGCAAGAATAAATCAGCTCAGGACACTACAAGATCGAGACTGTGCTCTTCATGCTCAGGAACGCATCTCGATTGGATTGCAT
 911 M E T V A V A R N I E N L S E A K I T P V L P A R L A S Q Q S K E T T
 2731 ATGGAGACCGTTCGCCGTTGCAAGAACATAGAGAACTTTCTGAAGCTAAAATTTACTCCTGTGCTGCTGCCCCCTGTCATCCCAGCAGTCTAAGGAACAATTT
 946 A S A G S G S K S S E R I H E Q E S S S H P M Q H A V S A R T D R Q W
 2836 GCATCTGCAGGATCCGGTCAAAGTGTGCAAGAAGATTCTGAGCAGGAAAAGCAGCAGCCATCCAAATGCAGCATGAGTTCTTCCGCAACTGCAGGCAAGTGG
 981 C V T V A S L A D A Q A C A K V T S Q N A G F I R N S P L Y K L I G E H
 2941 TGTGTCAGGGTGGCAGCTTAGCAGATCAAGCTGTGCTAAGTCACTCCAAAATGCTGGCTTTCATCAGAAACTCCCTCTGTCAAACTGATTGGGAAACAT
 1016 S V I A A V K P V S G E A V D K I E I E M H V G P D A A S K I V K T I
 3046 TCAGTCTAGTGTGTGAAACCTGTTTCCCGTTGAAGCCGTTGACAAGATAGAGATTGAAATGCATGTTGGAGCTGAGCAGCATCAAAAATCGTTAAAACCATC
 1051 T V K D D N A K E G H A G H S P V V L K L R E I L E T E K K ▽ **D H S R N**
 3151 ACAGTGAAGGACGAATAAAGGAGGCGCTCAGGACTCCCGGTTGTATTAAAATCGGAGAATCTGTAACCTGAAAAAGCAATTCAGGCAACTGAAAACGAACTGCAAGGAAC
 1086 **A T T S S S S S S S S S S R S R S Q K S S S S S S S S S S S S S S S**
 3256 GCCACCACTCCAGTTCAGCAGTTCTTCTAGCCGCTCAGGCCAGAGAGCAGTTCCCTCAAGTTCTTCTCCCTCCAGCTCCCTCAGCTCCTAGCAACAGC
 1121 **S K R S L K K N I K K R S**
 3361 AGTAAACGTAGCAAGAAAGAACATCAAGAAGAGAAGCAGCAGTACGAGCAGCAGCAGTGTAGTAGTAGCAGCAGCAAGAAAGCAGGACCCAAAATGT
 1156 L G G S L G S S S S S R S S R S S Q T S K A A I F Q R E T Q N H I H Q H E T
 3466 TTGGGGGAAGCAGTTCCAGCAGCAGCTGCTCTCCAGATCCAAAACATCCAAAGGGCCCATCTTTCAGAGGTTCCACAGAAATCACATATTATCAGCATGAAACC
 1191 **T R A A S S Q K T S S S P S S S S S S S S S S S S S S S S S S S**
 3571 ACTAGAGCTGCTCTTCGCAAAAGACCTCTCCTCAGCCCCAGTACGAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC

Figure 1b. Nucleotide sequence and deduced protein sequence of the N-terminal end of Vtg2 identified in *Heterotis niloticus* liver (N-ter2 HnVtg2). The start codon is in bold type. The signal peptide is underlined, lipovitellin I (LvI) is delineated by two black triangles, and partial phosvitin (Pv) is double-underlined.

1 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSGSSASSSFENIYK**KSR**
 1 AGCGGTAGCAGCGCCTCAAGCTTTGAGAACATTTACAAAAGTCTCGA
 36 FLGDTVRPAAVVIVRAVVRGNERRREGYQVAAYMDKA
 106 TFCCTTGGAGATACCGTTTCGTCGCCGTCGCCGTGCATCGTTTCGAGCCGTGAGAGGAAACGAGAGGAGCGGGGATACCAGGTTGCTGCCTACATGGACAAAGCT
 71 NAKRVQVILISALADSDKWLCLCAGLQLSMHKVMAKI
 211 AATGCCAGGGTGAGGTCATATAAGTGTCTTGTGACAGTGACAAATGGCAGCTGTGCCGTGATGGTATCCAGCTCAGATGCACAAAGTCATGGCCAAAT
 106 EWGAECQDYRAVIAKAFETGLLGLPHPAAOQLKMYWNKT
 316 CGTTGGGAGCAGAAATGCCAGGACTACAGAGCAGTTATCAAGCAGAGACTGGGCTTCTGGGTCACATCTCGACCCAACTGAAGATGTACTGGAACAAGACC
 141 FRALKRYASMITYEYVPGVALLAGFSEFRHRSGERO
 421 CCTCGTGCCTGAAGCGCTACGCTAGCATGATTTATGAATACGCCAGGTGTGCACCTTGGCTGGATTAGTGAAGGAAGACACAGAAGCGCGAGAGGCAG
 176 IKLTMAAATSAARITISILIRTPWMTLYLKLGQVIPIAL
 526 ATTAACTCACCATGGCTGCGACTTGTCCAGGACCATAGTATCATCCTTAGAACGCTTGGATGACTTTGTACAACTGGCCAGGTTATTCCTATTGCCCTG
 211 PIGAARAARAEEVEQTEAAGRIHYMFEVEAATSACKCKLEN
 631 CCCATGGAGCCGCTCCGCTCGTGGGAAGTGGAGCAGACTTTCGCTGGCAGGATCACTACATGTTTGTGAGGCCACTTCAGCCAAGTGCAAATGGAGAAC
 246 STVTTENNRE●RYGLQMP R S C Y Q V V A Q D C T S K L K F M V
 736 AGCACTGTACCACATTCACAACAAGGATGAGTGTCTTCAATGCCAGGTCCTGCTACCAAGTAGTGGCTCAGGACTGCACCTTAAGTCAAGTTCATGGTT
 281 L R K G D E R S E E S H V I V K I A D I D V D L T A E H G N I Q V K V
 841 CTGAGGAAGGGGATGAGAGATCTGAAGAATCCCACGTGATTGTGAAAATGGCCGACATTGATGTGGACCTGACTGCTGAAACCGTAAACATCCAGGTGAAAGTG
 316 N G R V V P I T Q H N Y E H P T G T I S I K Q K G A G I S L C A P S H
 946 AACGGAAGGGTAGTCCCTATCAACCCAGCACAATTATGAACATCCTACAGGTACCATCAATCAACAAAAGGCTGCTGGAATATCACTATGTGCTCCTCACAT
 351 G L H E V Y F N K N I L T I Q V P D W M K G N V C G L C G K A D G D V
 1051 GGTCTACACGAAGTCTATTTAAACAAGAACATCTTGACGATTCAGTACCAGACTGGATGAAAGGAAATGTTTGGGCTTTCGGGAAAAGCTGATGGTGACGTT
 386 R●REFEQGPSGHHEIDAVSFAHSWVLA AESCHDAKQC
 1156 AGAAGAGAATTTCAAGGGCCAGTGGACACCACATGAGGACGCTGTCACTTTTGCTCATTCTGGGTTCTTGTGCTGAAAGCTGCCATGATGCCAAGCAGTGC
 421 Q V K Q E L F K L T E P V L L N D Q D M K C Q S T F P V L S C L P Q C
 1261 CAAGTGAAGCAAGAGTTGGTTAAAGTGACGGAGCCAGTCTTCTGAATGACCAGGATGAAATGCAAATCTACTTTCTGTGCTGCTGCTCCACAAATG
 456 S P M K T T P V T V G F H C I P I D S N V N S W S S I R K K Q E D I R
 1366 TCCCAATGAAGACAAACCCTGTAAACAGTAGGGTCCACTGTATCCCCATGATTTCAATGTCAACAGTTGGAGCAGCATCAGGAAGAAGCAAGAAGCATCAGG
 491 V T V D A H I M C N C G E E C A *
 1471 GTGACC GTTGTGCTCACATATGTCAATTTGGGAGAAGATGTGCTTAA

Figure 1c. Nucleotide sequence and deduced protein sequence of the C-terminal end (C-ter) of Vtg1 and Vtg2 identified in *Heterotis niloticus* liver. The stop codon is in bold type. The phosphitin (Pv) is double-underlined, lipovitellin II (LvII) is in boxes, the β³-C domain is delineated by two black circles, and followed by the C-terminal (CT) coding region.

1 M R A V L L A L T L A L A A C▼**Q Q D S L T P D F A N G K A Y M F K Y E**
1 ATGAGAGCAGTCCTTCTTGCAGTACTAGCCCTTGCGGGGTCAACAGGACAGTTTAAACACCTGACTTTGCCAATGGGAAGCCCTACATGTTCAAATATGAA
36 **A E L L G G L P V E G L G K A G V K I V C K V L I S R V S Q N T Y L L**
106 **CGAGAGTACTGGGAGGTTTGGCCAGTGGAAAGGCTGGGCAAGGCTGGAGTGAAGATTGTCTGTAAAGTTCTCATCAGTCGAGTTTACAGAAACACCTATCTGCCTC**
71 **K L K E P Q I F E Y S G I R P R D D F K P A A K L T Q A L A A Q L L I**
211 **AGGCTCAAGGAACCTCAGATCTTTGAGTACAGCGGAATACGGCCAGAGAGATTCAAACCTGTGCAAACTCACCCAGGCCCTGGCCGCTCAGCTTCTTATT**
106 **P V K F E Y I R G V V G K V F A P A A V S K T I L N L H R G I L N I L**
316 **CCAGTCAAGTTTGAATACATCAGGGGTGTGGTGGCAAGTGTGGCCCTGCCCGAGTCCAAAACCTTCTCAACCTCCACAGAGGAATTTCAACATTTCTC**
141 **H V N I K K T Q N V Y E L Q E A G S Q G V C K T D Y A I S E D T K A E**
421 **CATGTCAACATCAAAAAGACACAGAATGTTACGAGTTGCAGGAGGCTGGATCCCAGGAGTCTGCAAACTGACTACGCAATCAGTGAGGATACCAAGGCTGAA**
176 **H I Y V T K S K D L G N C Q N R I M A D I G I A Y T E T C V Q C Q M R**
526 **CACATCTATGTGACCAAGTCCAAGGATCTGGGAACTGCCAGAACAGGATCATGGCAGACATTGGAATAGCTTACACAGAGACTTGTGTCAGTGCCAGATGAGA**
211 **N K N L R G A A T Y S Y V M K P T E T G A L I T K A T V Q E V H Q F T**
631 **AACAAGATCTGAGAGGAGCTGGCACCTACAGCTACGTCATGAAGCCACCGAAACAGTGTCTTATCACAAGGCCACAGTCCAGGAGGTTTATCAGTTCACT**
246 **P F H E L R G A G Q V E T R Q E M T F V E T Q N D P V Q P A H A D Y V**
736 **CTTTCCACGAGCTGAGAGGAGCGGGCCAGTAGAAAACAGAAAGAGATGACTTTTGTGAAACCCAGAAATGACCCAGTCAACCTGCACAGCAGACTACGTTG**
281 **A R G S L Q Y E F A S E L L Q T P L P L M K I T D I K T Q I E E I L N**
841 **GCAGTGGATCCCTCGCATGATGAATTTGCATCTGAGCTTCTCCAGACACCGCTGCCTCTGATGAAATCACTGATATCAAGACACAGATTGAAGAGATCTTGAAT**
316 **H L V K N N M G E V H E D A P L K F L Q L T Q L L R E A K Y E I M N G**
946 **CATCTGGTCAAAAACACATGGTGTGAGTTTACGAAAGTGTCCGCTAAAATTTCTTCCAGTACTCACTCCTCGGTGAGGCCAAAATATGAGATCATGAATGGC**
351 **I W T Q V K S K P V F R R W F L D T V P A V G N Q D A V R F I K E K F**
1051 **ATTTGGACTCAAGTCAAATCAAAACAGTATTACAGCGTGGTTCCTGGATACGGTTCCTGCGTGGAAATCAAGACGCTGTGAGGTTATCAAGGAGAAGTTT**
306 **I A G D I S A A E T A Q A L L V A L H L L E A N M D T V N L A G T L V**
1156 **ATAGCTGGTATATCTCTGCAGCAGAACTGCCAGGCTCTACTGGTTCAGCTGCATCTGCTAGAAGCTAACATGGACACAGTCAACTTAGCTGGCACTCTGGTA**
421 **F H A K L Q S H P M L R E I A M L G Y G S L V F K L C T K Q Q N C P A**
1261 **TTCCAGCCAAACTTCAGTCCCACCCATGTTACGTGAAATAGCCATGCTGGTTATGGTTCCTTGGTGTCAAGTTGTGCACTAAACACAGAACTGCCCGCT**
456 **E V I K P V H D I A A E A I S K A N V E E I A L V L K V L G N A G H P**
1366 **GAAGTCATAAAACCTGTTACGATATTGCTGCAGAAAGCAATTAGTAAGGGCAATGTGAAGAGATTGCCCTGGTCTGAAGTTCTGGGAAATGCAGGCCACCCG**
491 **A S I K P I M K V L P G F G S T A S S L T V K V H V D A V M A L R H I**
1471 **GCCAGCATCAAGCCGATCATGAAGTCTGCCTGGATTGGGTCTACAGCCTCTCACTTACAGTGAAGTCCAGTGTGATGCGGTGATGGCTTTGAGACACATT**
526 **A K R E N H K V Q E V A L Q L F M N R D L H P E V R M V A C V V L F E**
1576 **GCCAAAAGGAAACCAAAAGTCCAGGAAGTTGCCCTGCAGTTGTCATGAACAGAGATCTCCACCAGAAGTCCGCAAGTGGCTGCTGTGTGTGTTGTTGAG**
561 **C K P P I G L V A M I A S A L Q N E K S L Q V A S F A Y S H M K A L T**
1681 **TGCAAACTCCATCGGTCTGATGATTGCAAGTGTCTACAGAAGCAAGCAAGTGTGCAAGTGGCCAGCTTTGCCTATTCTCATGAAGGCCTTGACC**
596 **R S T A P E L L Q V A A A C N V A I K I L S P K L D R L S Y R F S K T**
1786 **AGGAGCACCGCCAGAACTTGCCCAAGTTGCAAGTGTGCCATCAAGATCTTGAGCCCAAGTTAGACTGACTGAGCTATCGCTTCAGCAAAACC**
631 **I H M D F F N Y H L M A G A A A T A H L I N D A A S I L P R A L V A K**
1891 **ATCCACATGGACTTTTTAACTATCACCTGATGGCTGGCCGCTGCAGCTATGCCATTTGATCAATGACGCTGCCAGCATTTTGGCAAGGCGCTTTGGCCAAA**
666 **I R A C M L G A A A D V L E F G V R T E G L Q E A L L K S P A A D P N**
1996 **ATCCGTCCTGCATGCTTGGGCGAGCTGCAGATGTCTGGAGTTGGTGTGAGAAGTGAAGACTCCAGGAGGCTCTTTGAAAAGTCCAGCTGCGGATCTTAAT**
701 **A C R G L T R M N R V L N T L K N N W K S L P A S Q P V A S A Y L K W F G**
2101 **GCTGACAGCTCACCCGGATGGCGTGTCTAAACACTCTGAAAACCTGGAAGTCACTACCAGCAAGCCAACAGTGGCTTCTGCTTACTTGAATGGTTGGC**
736 **Q E I A F A N I D R D I I E K A V E F A T G A A A Q P A L L K N I L N**
2206 **CAAGAAATGCTTTTGGCAACATTGACAGGACATCATCAGAAAGCCGTCGAGTTGCGCACAGGAGCTGCTGCACAGCCAGCTGTTGAAGAATCTCCTGAAT**
771 **M M Q S G I D I Q I A K P L M T S E V R R I E P T S M G F P I E I S L**
2311 **ATGATGACGCTGGTATTGATATCCAATTTGCCAAACCACTGATGACCTCTGAGGTGGCTGCGATCTTCCCCAGTCTATGGGGTTCCCATAGAGATCAGCCTC**
806 **Y S A A V A A A I V K A K A I L N P K P S D N F R I A Q L M N T D I Q**
2416 **TACTCAGCTGCTGTGGCTGCAGCTATAGTTAAAGCTAAAGCAATTCTGAATCCAAAACCTCCGACAACCTCAGGATTGCTCAGTTGATGAACACCCAGATTCAG**
841 **L D T R V V P S I A V H K Y A V M G V N T A L I Q A A I E A K V K V Q**
2521 **TTGGACACCCGTTGTACCAAGCATAGCTGTACATAAATATGCAGTTATGGGAGTAAACACTGCACCTATTGAGGCTGCAATTTGAGGCCAAAGTCAAAGTTGAG**
876 **K V L P L K F N A R I N I A Q G H Y K I E T V P L H A Q E R I L D I**▼**Q**
2626 **AAAGTCTTCTCTCAAGTTCAACCAAGAAATAAACATCGCTCAGGACACTACAAGATCGAGACTGTGCCTCTTATGCTCAGGAACCGCATTCCTGATTGCAA**
911 **I S C R C S Q E C R K S E R S**↓
2731 **ATAAGCTGTGCTGTAGTCAGGAAGTAGAAAAATCTTCCGAAGCTAA**

Figure 1d. Nucleotide sequence and deduced protein sequence of the Vtg3 identified in *Heterotis niloticus* liver. The start codon and the stop codon are in bold type. The signal peptide is underlined, lipovitellin I (LvI) is delineated by two black triangles, and LvII is in boxes.

```

1▼R V V P S I A V H K Y A V M G V N T A L I Q A A I E A K V K V Q K V L
1 CGTGTGTACCAAGCATAGCTGTACATAAAATATGCAGTTATGGGAGTAAACACTGCCTCATTGAGGCTGCAATGAGGCCAAGTCRAAGTTCAGAAAAGTCCTT
36 P L K F N A R I N I A Q G H Y K I E T V P L H A Q E R I L D L H M E T
106 OCTCTCAAGTTCACCGCANGAATAAACATCGCTCAGGGACACTACAAGATCGAGACTGTGCCCTTCATGCTCAGGAACGCATTCTGGATTGCATATGGAGCC
71 V A V A R N I E N L S E A K I T P V L P A R L A S Q Q S K E T F A S A
211 GTTGCCGTGGCAAGAACATAGAGAATCTTTCTGAAGCTAAAATTACTCCTGTGCTGCTGCCCGCCTTGCATCCAGCAGTCTAAGGAAACATTTGCATCTGCA
106 G S G S K S S E R I H E Q E S S S H P M Q H A V S A R T D R Q W C V T
316 GGATCCGGGTCAAAGTCGTACAGAAAGATTTCATGAGCAGGAAAGCAGCCATCCAATGCAGCATGCAGTTTCTGCCAGAACTGACAGGCAATGGTGTGTCCAG
141 V A S L A D Q A C A K V T S Q N A G F I R N S P L Y K L I G E H S V I
421 GTGGCAAGCTTAGCAGATCAAGCTTGTGCTAAGGTCACCTCCCAAAAATGCTGGCTTCATCAGAAACTCCCTCTGTACAACTGATTTGGAGAACATTCAGTCATT
176 A A V K P V S G E A V D K I E I E M H V G P D A A S K I V K T I T V K
526 GCTGCTG7GAAACCTGTTTCCGGTGAAGCCGTTGACAAGATAGAGATTGAAAATGCATGTTGGACCTGACGCAGCATCAAAAATCGTTAAAACCATCACAGTGAAG
211 D D N A K E G H A G H S P V V L K L R E I L E T E K K Q H S R N A T T
631 GACGACAATGCCAAGGAGGGCCATGCAGSACACTCCCGGTTGTATTAACACTGCGAGAAATCCTTAAACTGAAAAAACAGCATTCAAGGAACGCCACCACC
246 S S S S S S S S R R S G Q K S S S S S S S S S S S S S S S S S S S N S S K R
736 TCCAGTTC AAGCAGTTC TCTAGCCGTCGCTCAGGCCAGAAAGCAGTTCCTCAAGTTC TCTTCCCTCCAGCTCCTCATCCAGCTCTAGCAACAGCAGTAAACGT
281 S K K N I K K R S S S S S S S S S S D E G P D D H P I K K T K
841 AGCAAGAAGAACATCAAGAAGAGAGCAGCAGTAGCAGCAGCAGCAGCTCAGATGAGGGTCTTGATGACCACCCCTATAAAGAAGACAAA

```

Figure 1e. Nucleotide sequence and deduced protein sequence of the partial ovarian vitellogenin identified in *Heterotis niloticus* (HnVtg4). The lipovitellin I (LvI) is delineated by two black triangles, and phosvitin (Pv) is double-underlined.

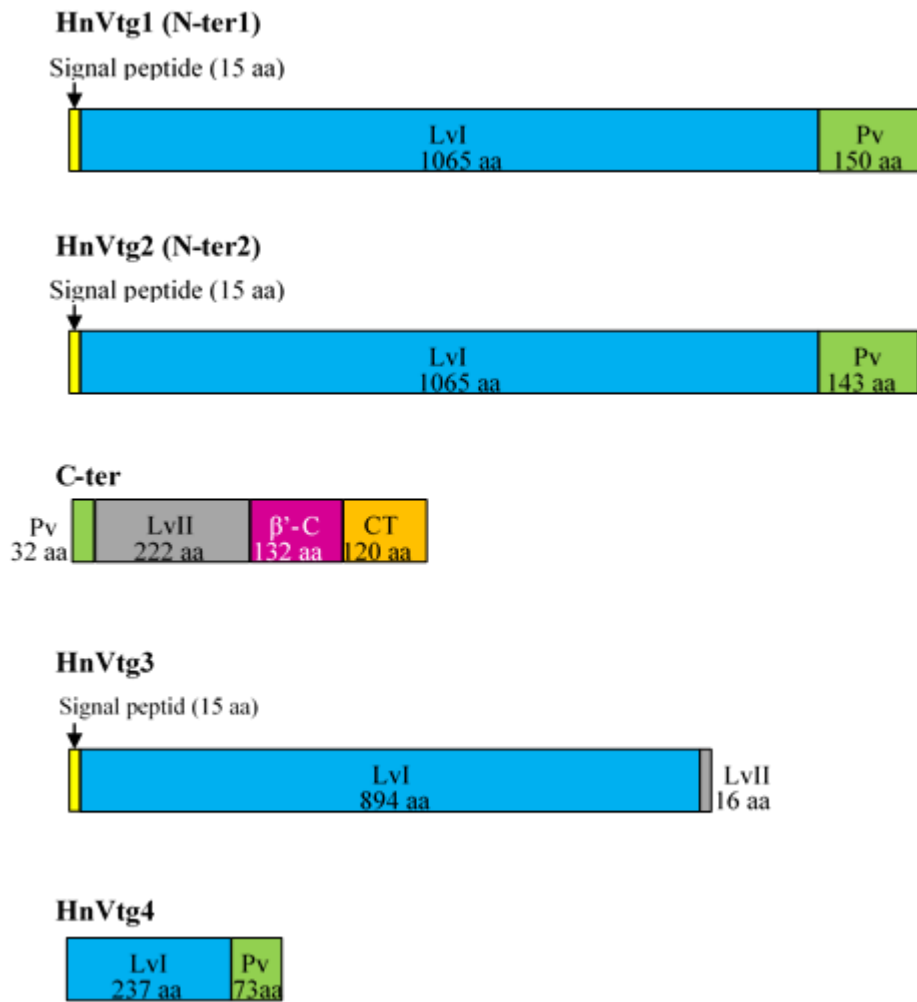


Figure 2. Linear constructions of the different *Heterotis niloticus* Vtg domains. The numbers of amino acids (aa) are indicated.

```

HnVtg1 (N-ter1) MRAYVLLALTLALAGOODSLTPDFANGKAYMFKYEAELGGLPVEGLBKAGVKIVCKVLSRVSQNTYLLKKEQIIFEYSGIRPRDDFK 90
HnVtg2 (N-ter2) MKAAVFALALALVAGQONILTPDFAAKKTYVYKYEAGLHSELPEEGLAKAGLKIASKVLSRADQNTYLLKKEQIIFEYSGIRPRDDFK 90
HnVtg1 (N-ter1) PAAKLTOALAAQLLIPVKFEYIRGVVGVVFAPAASVKTILNLHRGILNLIHVNIAKKTQNVYELQEAGSOGVCKTDYASIEDTKAEHIYVT 180
HnVtg2 (N-ter2) PAAKLTOALAAQLLIPVKFEYIRGVVGVVFAPAASVKTILNLHRGILNLIHVNIAKKTQNVYELQEAGSOGVCKTDYASIEDTKAEHIYVT 180
HnVtg1 (N-ter1) KSKDLGNCONRIMADIGIAYTETCVQCQMRNKNLRGAATYSYVMKPTETGALITKATVQEVHQFPPFHELIRGAGQVETROEMTFVETOND 270
HnVtg2 (N-ter2) KSKDLGNCONRIMADIGIAYTETCVQCQMRNKNLRGAATYSYVMKPTETGALITKATVQEVHQFPPFHELIRGAGQVETROEMTFVETOND 270
HnVtg1 (N-ter1) PVOPAHADYVARGSLQYEFASELLOTPLPLMKITDIKTOIEEILNHLVKNMGEVHEDAPLKFLQLTOLLREAKYIIMNGIWTQVSKSPV 360
HnVtg2 (N-ter2) PVOPAHADYVARGSLQYEFASELLOTPLPLMKITDIKTOIEEILNHLVKNMGEVHEDAPLKFLQLTOLLREAKYIIMNGIWTQVSKSPV 360
HnVtg1 (N-ter1) FRRWFLDTVPVAVGNQDVRFIKEKFIAGDISAAETAQALLVALHLLANMDTVNLAGTLVFHAKLQSHPLREIAMLGYGSLVFKLCTKO 450
HnVtg2 (N-ter2) FRRWFLDTVPVAVGNQDVRFIKEKFIAGDISAAETAQALLVALHLLANMDTVNLAGTLVFHAKLQSHPLREIAMLGYGSLVFKLCTKO 450
HnVtg1 (N-ter1) QNCPAEVFKPVHDAEAIISKANVEEIALVLKVLGNAGHPASIKPIMKVLPGFGSTASSLTVKVHVDAVMALRHIKRENHKKVQVEVALQL 540
HnVtg2 (N-ter2) QNCPAEVFKPVHDAEAIISKANVEEIALVLKVLGNAGHPASIKPIMKVLPGFGSTASSLTVKVHVDAVMALRHIKRENHKKVQVEVALQL 540
HnVtg1 (N-ter1) FMNRDLHPEVRMVACVVLFECKPPIGLVAMIASALONEKSLQVASFAYSHMKALTRSTAPELAQVAAACNVAIKILSPKLDRLSYRFSKT 630
HnVtg2 (N-ter2) FMNRDLHPEVRMVACVVLFECKPPIGLVAMIASALONEKSLQVASFAYSHMKALTRSTAPELAQVAAACNVAIKILSPKLDRLSYRFSKT 630
HnVtg1 (N-ter1) IHMDFFNHYLMAGAAATAHLINDAASILPRALVAKIRACMLGAAADVLEFGVRTGLEALLKSPAADPNADRLTRMWRVNLTKNWKSL 720
HnVtg2 (N-ter2) IHMDFFNHYLMAGAAATAHLINDAASILPRALVAKIRACMLGAAADVLEFGVRTGLEALLKSPAADPNADRLTRMWRVNLTKNWKSL 720
HnVtg1 (N-ter1) PASQPVASAYLKWFGQEIFAFANIDRDIIEKAVEFATGAAAOQPALLKNILNMQSGIDIQIAKPLMTSEVRRIFPTSMGFFIEISLYSAAV 810
HnVtg2 (N-ter2) PASQPVASAYLKWFGQEIFAFANIDRDIIEKAVEFATGAAAOQPALLKNILNMQSGIDIQIAKPLMTSEVRRIFPTSMGFFIEISLYSAAV 810
HnVtg1 (N-ter1) AAAIVKAKAILNPKPSDNFRIAQLMNTDIQLDTRVVPSIAVHKYAVMGVNTALIQAAIEAKVKVQKVLPLKFNARINIAQGHYKIEVPL 900
HnVtg2 (N-ter2) AAAIVKAKAILNPKPSDNFRIAQLMNTDIQLDTRVVPSIAVHKYAVMGVNTALIQAAIEAKVKVQKVLPLKFNARINIAQGHYKIEVPL 900
HnVtg1 (N-ter1) HAQERILDLHMETVAVARNIENLSEAKITPVLPARLASQSKETFASAGSGSKSSERIEHQESSHPMHAVSARTDRQWCVTVASLADQ 990
HnVtg2 (N-ter2) HAQERILDLHMETVAVARNIENLSEAKITPVLPARLASQSKETFASAGSGSKSSERIEHQESSHPMHAVSARTDRQWCVTVASLADQ 990
HnVtg1 (N-ter1) ACAKVTSONAGFIRNSPLYKIGEHSVIAAVKPVSGEAVDKIEIEMHVGPDAASKIVKTIIVKDDNAKEGHAGHSPVVLKLEILETEKK 1080
HnVtg2 (N-ter2) ACAKVTSONAGFIRNSPLYKIGEHSVIAAVKPVSGEAVDKIEIEMHVGPDAASKIVKTIIVKDDNAKEGHAGHSPVVLKLEILETEKK 1080
HnVtg1 (N-ter1) QHSRNATTSRSSSSSSRRSGOKSSSSSSSSSSSSSSSSSSSSSSSKRSKKNIKKRSSSSSSSSSSSSSSSSSSSSSSRRSRRTENVLGGSSSSSSRSRSQ 1170
HnVtg2 (N-ter2) QHSRNATTSRSSSSSSRRSGOKSSSSSSSSSSSSSSSSSSSSSSSKRSKKNIKKRSSSSSSSSSSSSSSSSSSSSSSRRSRRTENVLGGSSSSSSRSRSQ 1170
HnVtg1 (N-ter1) TSKAAIFORFTQNHIOHETTRAASSQKTSPPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1230
HnVtg2 (N-ter2) TSKAAIFORFTQNHIOHETTRAASSQKTSPPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1223

```

Figure 3a. Alignments of protein sequences deduced from the N-terminal end (N-ter1 and N-ter2) of the hepatic Vtg of *Heterotis niloticus* (HnVtg1 and HnVtg2) with little conserved domains in gray.


```

HnVtg4   ---RVVPSI AVHKYAVMGVNTAL I QAA I EAKVKVQKVLPLKFNAR IN I AQGHYK I ETVPLHAQER I LDLMETVAVARN I ENLS 81
HnVtg1 (N-ter1) LDTRVVPSI AVHKYAVMGVNTAL I QAA I EAKVKVQKVLPLKFNAR IN I AQGHYK I ETVPLHAQER I LDLMETVAVARN I ENLS 924
HnVtg4   EAK I TPVLPARLASQQSKET FASAGSGSKSSER I HEQESSHMQHAVSARTDRQWC V TVASLADQACAKVTSONAGF I RNSPL 165
HnVtg1 (N-ter1) EAK I TPVLPARLASQQSKET FASAGSGSKSSER I HEQESSHMQHAVSARTDRQWC V TVASLADQACAKVTSONAGF I RNSPL 1008
HnVtg4   YKL I GEHSV I AAVKPVSGEAVDK I E IEMHVGPDAAASK I VKT I TVKDDNAKEGHAGHSPVVLKLR E I LETEKKQHSRNATTSSSS 249
HnVtg1 (N-ter1) YKL I GEHSV I AAVKPVSGEAVDK I E IEMHVGPDAAASK I VKT I TVKDDNAKEGHAGHSPVVLKLR E I LETEKKQHSRNATTSSSS 1092
HnVtg4   SSSRRSGQKSSSSSSSSSSSSSSSSSSSSNSKRSKKN I KKRSSSSSSSSSSDEGPDDHP I KKTK----- 310
HnVtg1 (N-ter1) SSSRRSGQKSSSSSSSSSSSSSSSSSSSSNSKRSKKN I KKRSSSSSSSSSSSSSSSSRRSRTE NVLGGSSSSSSSRSSRSGTSKAAI 1176

```

Figure 3b. Alignments of protein sequences deduced from the N-terminal end (N-ter1) of hepatic Vtg (HnVtg1) and partial ovarian Vtg (HnVtg4) of *Heterotis niloticus* with little conserved domains in gray.

```

HnVtg1 (N-ter1) MRAVLLALTLALAGQODSLTPDFANGKAYMFKYEAE LGGLPVEGLGRAGVKIVCKVLISRVSONTYLLKLEKPOLFEYSGIRPRDDFK 90
SfMg (N-ter) MKAVVLLALTLALVAGQONNLTPDFATGKTYVYQYEALLOGGLPKEGLARAGVKIVSKVLISQVACTHLLKLEKPOLFOYTGWVPRDEFS 90
HnVtg1 (N-ter1) PAAKLTQALAAQLLIPVKFEYIRGVVVKVFAPAAVSKTILNLHRGILNLIHVNIKKTONVYELQEAGSQGVCKTDYVISEDTKAEHIVYV 180
SfMg (N-ter) PAAKLTQALAAQFRIPVKFEYVSGVVVKLFAPAAVSETILNLHRGILNLIQLNIIKKTQNVYELQEAGSQGVCKTDYVISEDTKAEHIVYV 180
HnVtg1 (N-ter1) KSKDLGNCQNRIMADIGIAYTETCVQCOMRKNLNRGAATYSYVMKPTETGALITKATVQEVHQFTPFHELFGAGOVETROEMTFVETOND 270
SfMg (N-ter) KSKDLGNCQKRYMADIGMAYTETCVQCOOKSKNLRGAATYSYVMKPTDSGALIMEAAVQELHQFTPFHELFGAQMKAQRLMTFVEAOND 270
HnVtg1 (N-ter1) PVQPAHADYVARGSLQYEFASELLQTP LPLMKITDIKTOIEEILNHLVKNNMGEVHEDAPLKFLOLQQLLREAKYETMNGIWTQVKS KPV 360
SfMg (N-ter) PVQPMGADYLARGSLQYEFASELLQTP LQLIKVTNAQAQIEEILNHLVKNNMGEVHEDAPLKFVELAQLLRMAKYETINKIWAQVKA KPD 360
HnVtg1 (N-ter1) FRRWFLDTVPVAGNODAVRFIKEKFIAGDISAAETAQALLVALHLEANMDTVNLAGTLVFHAKIQSHPLREIAMLGYGSLVFKLCTK 450
SfMg (N-ter) FRRWFLDTVPAIGTOVALRFIKEKFLAGEVTVIETAQALLAALHLVEANLDTVNLAAASVVLNAKIQSHPIREIAMLGYGSLVFKFCTE 450
HnVtg1 (N-ter1) QNCPAEVVKPVHDTAAEAISKANVVEIALVVKVGNAGHPASIKPIMKVLPGFGSTASSLTVKVHVDVAVMALRHIAKRENHKVOEVALQL 540
SfMg (N-ter) RNCPADVVKPIHDFAAEAISKANVAEIALAMKVLGNAGHPASIKPIMKLLPGFGSTAAALPVKVQVDVAVVALRHIAKREQRVODIALQL 540
HnVtg1 (N-ter1) FMNRDLHPEVRMVACVVLFECKPFI GLVAMIASALONEKSLQVASFAYSHMKALTRSTAPELAQVAAACNVAIKILSPKLDRLSYRFSK 630
SfMg (N-ter) FLDRDLHPELRMAACVVLFKTKPSI GLVSTIAAALQKEKSLQVASFAYSHMKALTRSTAPDLAQVAAACNVAIKILSPKLDRLSYRFSK 630
HnVtg1 (N-ter1) IHMDFFNHYHLMAGAAATAHLINDAASILPRALVAKTRACMLGAAADVLEFGVRTEGLQEALIKSFAADFNADRLTRMWRVNLTLKNWKS 720
SfMg (N-ter) IHLDFFNHBLMAGAAATAYEINDAATILPRAVVAKVRAYMVGAAADVLEIGVRTEGLQEALMKERAADAGADRLSRMRILLNALTNWKS 720
HnVtg1 (N-ter1) PASQPVASAYLKWFGQEIFAFANIDRDIIEKAVEFATGAAAOPALLKNI LNMMQSGIDTIQIAKPLMTSEVRRIFPPTSGFPTEISL YSAAV 810
SfMg (N-ter) PISQPLGSLVYLKFGQEIFAFANIDKDIIEBALQLATGAAAQHELWKTVLNLTQSGADFOISKSLLTSEVRRIFPPTSGFPMELSLYSAV 810
HnVtg1 (N-ter1) AAAIVKAKATLNPKPSDNFRIAQLMNTDIQLDTRVVPISAVHKI AVMGVNTALIQAAIEAKVYKQKVLPLKFNARINIAQGHYK IETVPL 900
SfMg (N-ter) AAAIVKAKATLTPPPRENFQLAQLKNTDIQLQAHIAPSIAVHKI AVMGVNTALIQAAIVAKAKVHNVLPKFNARVHIAQGHFKIEALPL 900
HnVtg1 (N-ter1) HADERILDLHMETVAVARNIENLSEAKITPVLPARLASQSKETFAS- - -AGSG-SKSSERIEQESS-HPMOHAVSARTDRQCWCVT 983
SfMg (N-ter) QAHGRLDLQMEAVAMSRNVENLSGAKITPVLPERLAALQSRERFTSRADADAGSGLSKSSERIEYENVSDEKRPKONDVSRMDK KWCVT 990
HnVtg1 (N-ter1) VASLADQACAKVTSQAGFVRNSPLYKLI GEHSVIAAVKPVSGEAVDKIEIEMHVGPDAASKIVKTIITVKDDNAKEGHAGHSPVVLKLR 1073
SfMg (N-ter) MANLGOQACAKVTSQAGFVRNSPLYKLI GEHSVTLQVVKPVSGEAVEKIEIEIQVGNAAASKIVKRIITVKDDNTDKGNRGGTPVVLK LK 1080
HnVtg1 (N-ter1) ILETETKQHSRNATSSSSSSSSRRSGOK- - -SSSSSSSSSSSSSSSSSSSSSSSKRSKKNIKKRSSSSSSSSSSSSSSSSRRRSRTE NVLGGSS 1159
SfMg (N-ter) ILETERTQNLKSSNSSSSSSSSSRRSGONR- - -SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSRK- - -VAGAWSNG 1167
HnVtg1 (N-ter1) SSSSSSSSSSSTSKAAIFORFTQNHITHQHETTRAASSOKTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1230
SfMg (N-ter) RSVSSSSSSSRISKAHVFRFTKXNHMHQDDNRATSORKISSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1238

```

Figure 3c. Alignment of the protein sequences deduced from the N-terminal end (N-ter1) of the Vtg1 of *Heterotis niloticus* (HnVtg1) and the N-terminal end -N-ter) of the Vtg of *S. formosus* (SfVtg) with little conserved domains in gray.

```

HnVtg1 (C-ter) SSSSSSSSSSSSSSSSSSSSSASSFEN I YKKS RFLGDTV RPAAVV I RAVRGNERRR GYQV AAYMDKANR VQV I T S A L A D S K W O L C 90
SfMg (C-ter) SSSSSSSSSSSSSSSSSSSSSASSFE A I YKKS RFLGDSV RPAAVV I I RAVRGNDRR GYQV AAYMDKAD R VQV I M A A L A E N D N W K M C 1329
HnVtg1 (C-ter) A D G I Q L S M H K V M A K I R W G A E C Q D Y R A V I K A E T G L L G P H P A Q L K M Y N K T P R A L K R Y A S M I Y E Y P G V A L L A G F S E G R H R S G E R Q I K L T M 180
SfMg (C-ter) A D G I Q L S M H K V M A K I R W G E E C Q D Y A T V I D A E T G L L G S H P A V Q M K L N W K K I P R A V K R Y A R M I S E Y I P G A A L M A G F S E G R H R S K R Q I K L T V 1419
HnVtg1 (C-ter) A A T S A R T I S I L R T P W M T Y K L G Q V I P I A L P G A A A A R A E V E C T F A G R I H Y M F E A T S A K C K L E N S T V T T F N N R R Y G L O M P F S C Y Q V Y A Q 270
SfMg (C-ter) A A T S A R T L S V I L K T P R M T M Y K L A Q A I P I A L P G A A A A R A E V E C N I A N R I H Y M F A E A N T A K C I V A D D T V T T F N N R R Y S P E V P S S C H O V L A Q 1509
HnVtg1 (C-ter) D C T S K L K F N V L R K G D E R S E E S H V I V K T A D I D V D L T A E H G N I Q V K V N G R V P I T O H N Y E H P T G T I S I K O K G A G T S L C A P S H G L H E V Y F N K N 360
SfMg (C-ter) D C T T E L K F V V L M K K D E S E K P H I V K V A D I D V D S A D D N G L H M K V N G M D V P T V K L P Y E H S T G S I I I R C N G D G L S L Y A P G Q G L H E V Y Y D G N 1599
HnVtg1 (C-ter) I L T I Q V P D W M K G N V C G L C G K A D G D V R R E F O G P S G H R I E D A V S F A H S W V L A A E S C H D A K C C O V K O E L F R L T E P V L L N D O D M K C Q S T F P V L S 450
SfMg (C-ter) I L R I E V E D W M K G N V C G L C G K A D G E V R D E Y Q A P S G R G I M S A V S F G H S W V L A A E S C H D A S E C R L R Q E S V M L D K P V T L H G O A S K C Y S T E P V L O 1689
HnVtg1 (C-ter) C L P D C S P M K T T P V T V G F H C I P I D S N V N S W S S I R K K C E D I R V T V D A H I M C N C G E E C A 506
SfMg (C-ter) C L P E C K P V K T T P V T V G F H C V P A G S R V N N L V S M N E K T E D L R D T L D A H V S C R C T E N C A 1745

```

Figure 3d. Alignment of the protein sequences deduced from the C-terminal end (C-ter) of the Vtg1 of *Heterotis niloticus* (HnVtg1) and of the Vtg of *S. formosus* (SfVtg) with little conserved domains in gray.


```

HnVtg3  MRA-----VLLALTLALAAAGOODSLTPDFANGKAYMFKYEAE LLGGLPVEGLGKAGVKIVCKVLIISVSONTY 68
OnVtgC  MTAHAYKRTALKPDLHSSLTMRGLVFFCCLVALATCOAARYDLSLNHRKTYEYTYEGEVKFGLGKPNRAKSGARIKCNIKISGESPOAF 90
OmVtgC  MWG-----FLLCHLVALAACAHISYEPGLNPKKTYEYTYEGVMNIGRGMPNLVESGVRLVCKVKIIGVSAQTF 68

HnVtg3  LLKLEKPEIFEYSGIRPRDDFKPAAKLTQALAAQLLIPVKFEYIRGVVGVKVFAPAAVSKTILNLHRGILNHLVNIKKTONTVYELQEAGS 158
OnVtgC  ILLVSDVTFEELNGISGKSDFNQSPKLTORLAAQLAKPFKFEHNSGHVROIQAAPVSETIVNIIVRGILSFFHATVKTSPTVYEELEAGI 180
OmVtgC  LLRVSNLNFEEFNGIPGKNAFDASPKLTKRIAAELSKPFMFEYAKRGVGNIRTSAEISDTIVNIIVRGILGFFQVTVKTTORTVYEELEE 158

HnVtg3  GGVCCKTDYATSEDTKAEHIVYTKSKDLGNCONRIMADIGIAYTETCVOCOMRNKNLPGAATYSYVMKPTETGALITKATVQEVHOFTPFH 248
OnVtgC  HGMCOQSNYATEINKOTNNMNI TOVVDINDNCRMKAAOMNGMASAVLDKDKSKORGESVISTVOYLVTIOPTEEGLLVKSAVGLQOHFSPFN 270
OmVtgC  HGLCOSTYANEEDANSQELSVTQITVDINCKKKAAYRGMALAVDNKLSKERGDSVISTVRYVYTKPTVDGGLIKKHAHALEROHFSPFN 248

HnVtg3  ELRGAGOVETRQEMTFVETQNDRVQPAHADYVARGSLQYEFASELQOTPLPLMKITDIKTGIEELNLHVKNMGEVHEDAPLKFLOLTO 338
OnVtgC  VKRGSFRMEAKKEMVLIRVSEGGKPANVQGLQNRGDLVYKLVNSDVNIPLVMODLAEPKAKAVRMIKQLAEDHKNQINRETTEDTLKVVYO 360
OmVtgC  VKRGNKSLGATKEELVLLSVTDTVTPAVGPMISRGNLVYKFKVLAQIPIPLMKKLDPPVRKIADMIKRLAQAANIYOTDSATSEDDVVVLF 338

HnVtg3  LLREAKYEIMNGIWTQVKSPPVRRWFLDTPAVGNQDAVRFIKEKFIAGDLSAETAQALLVALHLLLEANMDTVNLAGTLVHFHAKLQSH 428
OnVtgC  LLRMLQDRDLDEMVELARNDERRRFLDMIVEINDRVLKFLKRFQAKDVSANEAVOALLIAMNHLQPRAELEKAKEIMNMEFIKSN 450
OmVtgC  LLRVTSLNENHEVLWVKQLSGNDEHRRWLDITVEVADARVLKFLKRFKAGDVSANEAGOLLIAFNHLAETDLMEMAKEFLSMFPFKSH 428

HnVtg3  PMLREIAMLGYGSLVFKLCTKCO--ONCPAEVIKPVHDIAAEASKANVEEIALVLYKVLGNAGHPASIKPIMKVLPGFSTASSLTVKVHV 516
OnVtgC  LHMWPAAVSFGSLVYKYCAHY--SPCQSAVOPLDLANEGLRNRNETEMVIALRAGNAGHPSSIKTIANFLPGVAANPVDVSCRVS 538
OmVtgC  PMLWNTVLSYGSLLVRYCTEYDRCPCAAVVOPLDLVIDGLKRNNEFDMVVALKAIIGNAGHPSSIKTIIRRFPGVSAAPVLPARVLS 518

HnVtg3  DAVNALRHIAKRENHKVQEVVALQLFMNRDLHPEVRMVAQVLFCEKPPIGLVAMIASALONEKSLQVASFAYSHMKALTRSTAPELAQVA 606
OnVtgC  AAVQALRLTARRNPHNVOEIALKVFLOKNCAPETSMALIMVMFNTKPSAALVSTTAHLEREKDIHLASFAYSYLENIARSKTPDNHYS 628
OmVtgC  AAVQSLRHLAVRDPHVSODITVSLFVQRHLPTIIRMLACMLFETKPPALVSTVTAFLLEADMHVASFSYSLIGSIAKSTPDNHLS 608

HnVtg3  AACNVAIKILSPKLDRLSYRFSKTIHMDFFNYHLMAGAAATAHLINDAASILPRALVAKIIRACMLGAAADVLEFGVREGLQELALLKSPA 696
OnVtgC  IYCSLAVKSLAAKFGRLSYN-SRSLQVDMFDDLLMGTSASINVLGGPSTVLPTEIMSKNKYFAIGRIMRMLMEFIRAESLQKLFGPSIP 717
OmVtgC  TACNVAVKILAPKFGRVSYHYSKAMHLDWFNDDELIGTATEIFMLKNAASYIPTIEIVTKGKFHEIGKILQLVEIGFRADGIKELFRQNI 698

HnVtg3  ADPNADRLTRMVRVNLTLKNWKS LPASQPVASAYLKWFGQEIFAFANIDRDIIIEKAVEFATGAAAOPLLKNI LNMMQSGIDIOIAKPLMT 786
OnVtgC  GFKGDFSSDFQAIYNVLKNWESLPNDKPIMSAYSRSQGEIFFVNIKKDLLEYIKAAVSPSAPKGTPTVWAAIEELLQKGFSSHWHTKPLLS 807
OmVtgC  VFNHGLSYADFEAIFKVLQDWKTLPGDKPLLSAYTRIFGOEFFADLNNDLIGOSVIKTVSPSAGKESPVWSMIENLQKGMSSHWHTKPELM 788

HnVtg3  SEVRRIFRFTSMGEPTEISLYSAAVAAAIVKAKAILNPKPSDNFRIAQLMNTDIDLQTRVVPISAVHKEYAVMGVNTALLQAAIEAKVKVOK 876
OnVtgC  IEARLLQATLGLLPEIAKYYQTVGTITMNAKAAINPPOTD--RLGDLNADISLESEGVMGFTKDFWIFHGINTELFQCGSELKSKARV 895
OmVtgC  FETRYIQATSLGLPLEISKYYNTVTTITVNAKAAISPLTDD--RLGOLLTSDVSLVTDGFAGFSKQDFVFHGINTDIFQCGAELKSKTVV 876

HnVtg3  VLPKFNARINIAOGHYK IET----- 897
OnVtgC  ELPFKLSAKLVNGKLLLEDIPACKKEFELISVSSNVYAVSRNIGDPESPKRTPMIPD--ESNSQQOPTPNW-----HRKA 970
OmVtgC  AIPWEFSMKMNVKOKTTEIDLPAKKEETKLESVKFDVYAVSRNIEEPSRAKMTMMPDSEANKDIEPPKORYLTSQOKIVKFDIYHPQV 966

HnVtg3  -----VPLH----- 901
OnVtgC  NMC AKSS IYGVGVCDVDEIRROYFNNEYPLYSILGFTQAAIRVVP AEATKPVEKIHLEVNNGNPTSHPMTRHLYETMRRISK--ERSDEQ 1058
OmVtgC  KVC AEANIYGAVECVDEFETRSHYIDEVPLYFLGYTNFACTIQPVQSNKPVDKIHIEMNASLTKYPVSVRHLLDTLQKLPKYATKSFHH 1056

HnVtg3  -----AOERTLD----- 908
OnVtgC  SNSASNERETRNSQEIIMMDATAPO-----PVVSIKAIAMSGNONLEGFVTL YNPSEGDTLKAOMIVSHVGD AANWKMCDAS AQSPA 1141
OmVtgC  SPSSSSRSSHQASFKAKEASEALNANSEPVFAIKVLANMNGNTKREGYVAAYPTPAQMDNTOLISQIGEEANWKLCTDANVDKVR 1146

HnVtg3  ---LQISCRCSGECRK----- 921
OnVtgC  OSSANLAKAISWGAQCKPYEISVTA AAAQOSSASRP TLEAKVQWANVPEEMANYCSMERYIPGMALLSGFNQWESNANQEV SASIVSAS 1231
OmVtgC  QEQ--LKMHLRWAECQSYEMSMTAT AHLPGSRPTLKAJIHWKTVPGYMIEVGSIEKEYIPGMALLGFYQKHATNAKHEVSASVVAAT 1234

HnVtg3  -----SFRS----- 925
OnVtgC  SDSIDITIKLPKYTVFRCAVPLPFSSTSFRE-----ENAEERNGOA 1271
OmVtgC  AGSIATTIKLPDLTVYHCEIAIPIQVTDIFEDGOOLVGNSTLEMHKEL 1281

```

Figure 3f. Alignment between the VtgC of *Heterotis niloticus* (HnVtg3) and the VtgC of *O. mykiss* (OmVtgC) and *O. niloticus* (OnVtgC) with little conserved domains in gray.

1 M V T S I L G F L I L P I C L Q Q C A Y V R G S Q A E C D P S Q F Q C
 1 ATGGTCACGTCCTATTCTGGGCTTTCTTATCCCTGCCAATATGTCCTTCAGCAGTGTGCCTATGTCGGGGGATCCAGGCGGAATGTGATCCCTCTCAGTCCAGTGT
 36 G N G R C I P S V W Q C D G D E D C S D G S D E N T C V R K T C A E V
 106 GGAAATGGC CGCTGTATTCCCTCAGTGTGGCA GTGTGACGGGTGATGAA GACTGTTCCTGA TGGCAGTGATGA AAACA CCTGTG TGAGG AAGAC CTGCGCAGAAGT G
 71 D F V C R N G Q C V P K R W H C D G E P D C E D G S D E S S Q V C H M
 211 GACTTTGTC TGGCCG AATGGC CAGTGTGTCCCAARGA GGTGGC ACTGTGATGGCGAGCC CGATTG TGAAGA CGGGT CCGAGC AARGC TCCCA AGTCTG CCAACAT G
 106 R T C R M N E F S C G A G T T Q C I P V F W K C D G E K D C D N G E D
 316 CGGACCTGT CGTATG AACGAG TTCAG CTGTGG GGCAG GTACA ACCCAG TGCACT CCCAGT TTTCTG GAAAT GCGATG GAGAG AAGGA CTGCGA CAACGG GAGGAT
 141 E V N C G N I T C A P L E F T C A S G R C V S Q N F V C N G E D D C G
 421 GAGGTTAATCTGGCAACATC ACGT CGCCCC CTGGAGITCA CCTGT GCAAG CCGAGC CTGTGT GTCTCA GAATT TTTGTG TGCAT GAGCA GGCAGG A CTGCGGT
 176 D G S D E L D C E P S S C G P S Q F Q C G N A T C I P D S W V C D D
 526 GATG GCGAG CAGAG CTAGAT TGTGA GCCCTC ATCCT CGGGAC CCGAGT CCAGT GCGAG TGGCAA TGCCAC CTGTA TCCCT GACAGC TGGGT ATGTGA TGATGAT
 211 V D C Q D Q S D E S P Q R C G R H P T P P A K C P S S E M R C G S G E
 631 GTGGACTGC CAGGAC CAGTCA GATGA GTCAAC ACAGC GCTGTG GCCCG CACCC TACACC TCCCGC CAAGTG CCGTT CCGAGC AGATG CGCTG TGGCTC GGGCGAG
 246 C I H R K W R C D G D A D C K D G S D E D N C P V R T C R P D Q F K C
 736 TGCA TCCAC CGCAAG TGGCGT TGCGA TGGCGA TGCGG ACTGTA AGGAT GGCAG TGACGA GGCACA CTGCCC TGTCG GAACTT GCGA CCGA CCGATT CAGTGT
 281 N D G S C I H G S R Q C N G M R D C S D G S D E L N C K N V S Q C S G
 841 AATGATGGT AGCTGC ATCCAT GGCAG CCGGCA GTGCA ATGGGA TGCGT GACTG CTGAGA TGGCAG TGATGA GCTCA ACTGCA AAAAT GTGT CCAATG CAGTGGC
 316 P D K F K C R S G E C I E M S K V C N K V R D C S D W S D E P L K E C
 946 CCAGATAAA TTCAG TGGCGT AGCGG CGAGTG CATCG AGATGA GCAAA GTGTG CAACA GGTCCG GGAAGT CTCAG ACTGGA GCGAT GAGCC CCTCAAAGATGGC
 351 N L N E C L Q N N G G C S H I C R D N I I G Y E C D C T P G L Q L I D
 1051 AATC TGAAT GAATGC CTTTCAG AACAA TGGTGG CTGCT CCCACA TCTGT CGGGA CAACAT CAITGG TTACGA ATGCG ACTGCA CTCCA GGCCT CCGACT CAITGAC
 386 R K T C G G D I N E C L N P G I C S Q I C I N L K G G Y K C E C H N G Y
 1156 CGCAAGACC TGTGGT GATATT AATGA GTGCC TAAACC CTGGGA TCTGT AGCCA GATCTG CATCAA CTTGAA GGGGG GGTACA AGTGT GAGTG TCACAA TGGGTAC
 421 Q M D P T T G V C K A V G K E P C L I F T N R R D I R K L T L E R E
 1261 CAGA TGGAT CCTACC ACAGGC GTGTG CAARGC AGTGG GCAAGG AGCCA TGCC T GATCTT CACCAA CCGGCG TGACA TCCGCA AGCTG ACCCT GAGGCG GCGGGAG
 456 Y T Q I V E Q L R N T V A L D A D F T Q Q M I F W A D L G Q K A I F S
 1366 TACA CCGCAITGTG GAGCAG TTGCG CAACAC TGTGG CCTTA GATGCA GACTT CACCCA ACAGAT GATCT TCTGGG CTGAT CTAGG TCAAAA GGCAT CTTCAGC
 491 M S L D K R D E G S I T K V I E S V Q T P V G I A V D W I Y K N I Y W
 1471 ATGT CACTT GACAAG CCGGAC GAAGG CAGTAT CACAA AAGTGA TTGAG AGTGT GCGAGC TCTGT CGGGA TTGCA G TAGAG TGGAT TTACAA GAATAT ATACTGG
 526 S D L G T K T I S V A N F N G T K T K I L F S S G L K E P A S I A V D
 1576 TCAGATTGGG TACC AAGAC GATAT TGTAG CAACCT TCAATG GCAAGC AAGAC CAAAAT TTTGTT CAGCAG TGGGG TAAAG GAAAC GGCCTC CAITGCTGTGGAT
 561 P L S G F L Y W S D W G E P A R I E K S G M N G V D R Q V L V E T D I
 1681 CCTC TCTCT GGGTTT CTGTAT TGGT AGACTG GGGTG AGCCAG CCGAGG AITGA GAAGTC TGGCAT GAATGG TGTAG ATCGGC AGGTT CTGGT TGAGAC GGCAT T
 596 Q W P N G I T L D L I K S R L Y W V D S K L H M L C S V D L N G D N R
 1786 CAGT GGCCT AATGGG ATCACT CTAGA CTTGAT TAAAA GCAAGC TCTAC TGGGT GGAAGT CAAACT ACACAT GCTCT GCGAGC TGGAT CTGAA CCGGGG A TACCGG
 631 R K V L T Q S P D Y L A H P P A L T V F E D R V F W T D G E N E A I Y G
 1891 CGGA AAGIT CTCAG TCTCCT GACTA TCTCGC GCATC CTTTTG CTCTC ACGGT TTTTGA GGAATCG AGTGT CTGGA CAGATG GGGAA AARGA GGCCTC CTATGGT
 666 A N K P T G S D V I M L A S N L N E P Q D I I V Y H E L I Q L S G T N
 1996 GCCA CAAA TTCACT GGGTCA GATGT GATTAT GCTTG CCGGCA AACTG AATGA GCCACA GGCAT CAITGT TTATC ATGAGC TCATC CAGCT ATCCGG AACCAAT
 701 W C T D K G D N G G C A P L C L P A P Q I N K H S P K Y T C I C S Q G
 2101 TGGT GCATC GACAAA GGGGAC AATGG AGGCTG TGCC T TCTGT GCTTG CCGAGC ACCACA GATCAA CAAGCA CTCTC CAAAGT ACACCT TGCAT TTGCTC ACGGGG
 736 M E L A A D G L N C R P A T I S A H P G D G K T L I H P T H A S E G N
 2206 ATGG AGCTT GCTGCC GATGGC TTGAA CTGCGAG CACG CCAACCA TTTCT GCACA CCCAGG AGATGG AARGAC ACTAA TACACC CCACT CATGCTTCAGA GGGGAAT
 771 I S T S I H E V N S S A K G S A A A W A I L P V L L A M A V A G G Y
 2311 ATTAGTACC TCCATC CATGAG GTGAA CTCTC AGCCA AAGGGT CAGCT GCTCG ATGGGC CATCCT CCGTGT CTTG TGCTG CCAAT GCTGT GGCAGG AGGCTAT
 806 L M W R N W Q L K N K K S M N F D N P V Y L K T T E E D L N I D I S R
 2416 CTAATGTGG CGCAAC TGGCAG CTGAA GAACAA AARGA GCATG AACTTT GACAA TCTCTGT GTATCT GAAGAC CACCG AGGAG GACCTC AACAT CGACAT CAGCCGG
 841 H G A S V G H T Y P A I S V V N T E D D L *
 2521 CACG GTGCA TCTGTG GGCAC ACGTA CCTGCT TATAT CAGTTG TGAAT ACAGA AGATGA CTTGTAG

Figure 4a. Nucleotide sequence and deduced protein sequence of the *Heterotis niloticus* vitellogenin receptor or very-low-density lipoprotein receptor (HnVLDLR). The start codon and the stop codon are in bold type, the signal peptide is underlined, the eight repeats that bind to the ligand (LBR) are delineated by two consecutive black triangles, the three EGF1, EGF2 and EGF3 domains are located in the boxes, the highly conserved SDE, EDE, YWS(V)D, FWA(T)D and FDNPVY sequences are underlined, the O-glycosylation domain is double-underlined, and the transmembrane domain is delineated by two black circles and followed by the cytoplasmic domain in C-terminal position.

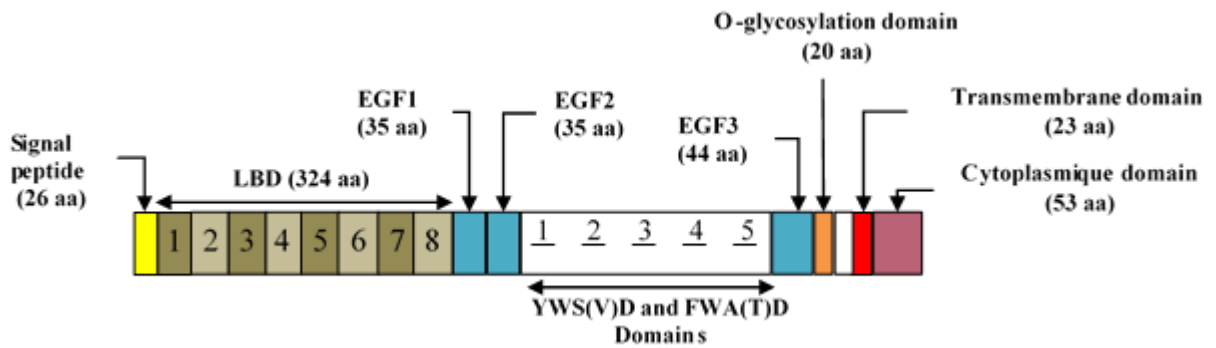


Figure 4b. Linear constructions of the different *Heterotis niloticus* Vtg receptor domains. The numbers of amino acids (aa) are indicated.


```

HnVLDLR MVTSLGLFILP--ICLQOCAYVWGSOAECDPSSQFCCNGRCHIPSVWCCDGDDECSGDSSENTCVRKTCAEVDFVCRNGCCVPKRWHCD 87
ObVLDLR MLASILGLFILP--ICLOHYVCVWGSOAECDPSSQFCCNGRCHIPSVWCCDGDDECSGDSSENTCVRKTCAEVDFVCRNGCCVPKRWHCD 87
SMDLR MVTSLGLFILP--ICLQOCAYVWGSOAECDPSSQFCCNGRCHIPSVWCCDGDDECSGDSSENTCVRKTCAEVDFVCRNGCCVPKRWHCD 87
OmVLDLR MLTSLLELILP--ICLQOCGFVHGSKECEPSSQFCCNGRCHIPSVWCCDGDDECSGDSSENTCVRKTCAEVDFVCRNGCCVPKRWHCD 87
CcVLDLR MVTSLGLFILP--VCLQOCGFVHGSKECEPSSQFCCNGRCHIPSVWCCDGDDECSGDSSENTCVRKTCAEVDFVCRNGCCVPKRWHCD 87
OaVLDLR MVTSTQGLLPLMLICLOHFVNVHGTKTECANQFCCNGRCHIPSVWCCDGDDECSGDSSENTCVRKTCAEVDFVCRNGCCVPKRWHCD 89
HnVLDLR GEPDCEGGSDSESSQVHMRTRRNNEFSCGAGTTCQIPVFWKCDGEEKCDNGEDEVNCGNITCAPLEFTCASGRFVSONFVNGEDDCGD 176
ObVLDLR GEPDCEGGSDSESPDVCYHTRTRCRVNEFSCGAGTTCQIPVFWKCDGEEKCDNGEDEANCGNITCAPLEFTCASGRFVSONFVNGEDDCGD 176
SMDLR GEPDCEGGSDSESPDVCYHTRTRCRVNEFSCGAGTTCQIPVFWKCDGEEKCDNGEDEANCGNITCAPLEFTCASGRFVSONFVNGEDDCGD 176
OmVLDLR GEPDCEGGSDSERVEVCHTRTRCRVNEFSCGAGTTCQIPVFWKCDGEEKCDHGEDMSCGNITCAPLEFTCASGRFVSONFVNGEDDCGD 176
CcVLDLR GEPDCEGGSDSETMEMCHTRTRCRVNEFSCGAGTTCQIPVFWKCDGEEKCDNGEDEINCGNITCAPLEFTCASGRFVSONFVNGEDDCGD 176
OaVLDLR GEPDCEGGSDSESLCHMRTRRNNEFSCGAGTTCQIPVFWKCDGEEKCDNGEDEVNCGNITCAPLEFTCASGRFVSONFVNGEDDCGD 178
HnVLDLR GSDELDCEPSSCGPSQFCCGNATCIPDSWVCCDDVDCDDQSDSPQRCGRHPPTPAKCSSEMFCSGSECIHRKWFCDGSDADCKDGSDE 265
ObVLDLR GSDELDCEPSSCGPSQFCCGNATCIPDSWVCCDDVDCDDQSDSPQRCGRHPPTPAKCSSEMFCSGSECIHRKWFCDGSDADCKDGSDE 265
SMDLR GSDELDCEPSSCGPSQFCCGNATCIPDSWVCCDDVDCDDQSDSPQRCGRHPPTPAKCSSEMFCSGSECIHRKWFCDGSDADCKDGSDE 265
OmVLDLR GSDELDCEPSSCGPSQFCCGNATCIPGNVCCDDVDCDDQSDSPQRCGRHPPTPAKCSSEMFCSGSECIHRKWFCDGSDADCKDGSDE 265
CcVLDLR GSDELDCEPSSCGPSQFCCGNATCIPPSWVCCDDVDCDDQSDSPQRCGRHPPTPAKCSSEMFCSGSECIHRKWFCDGSDADCKDGSDE 265
OaVLDLR GSDELDCEPSSCGPSQFCCGNATCIPASWVCCDDVDCDDQSDSPQRCGRHPPTPAKCSSEMFCSGSECIHRKWFCDGSDADCKDGSDE 267
HnVLDLR DNPVPTCRPDQFKCDGSCIHGSRCCNGLRDCSDGDELNCKNVSQSGPKFKCRSGECIEMSKVCKNVRDCPDWSDEPLKECNLINE 354
ObVLDLR ANCPVPTCRPDQFKCDGSCIHGSRCCNGLRDCSDGDELNCKNLSQSSPKFKCRSGECIEMSKVCKNVRDCPDWSDEPLKECNLINE 354
SMDLR ANCPVPTCRPDQFKCDGSCIHGSRCCNGLRDCSDGDELNCKNMTQSSPKFKCRSGECIEMSKVCKNVRDCPDWSDEPLKECNLINE 354
OmVLDLR ANCSVPTCRPDQFKCDGSCIHGSRCCNGLRDCADGTEVNCNKLTCOONGPKFKCRSGECIEMSKVCKNVRDCPDWSDEPLKECNLINE 354
CcVLDLR KNCVARTCRPDQFKCDGSCIHGSRCCNGLRDCADGTEVNCNKLTCOONGPKFKCRSGECIEMSKVCKNVRDCPDWSDEPLKECNLINE 354
OaVLDLR ANCPVPTCRPDQFKCDGSCIHGSRCCNGLRDCADGDEANCNMTQNGPKFKCRSGECIEMSKVCKNVRDCPDWSDEPLKECNLINE 356
HnVLDLR CLQNNGGCSHICDNDVIGYECDDTPGLQLIDRKTCDGIDNECLNPGICSOICINLKGKGYKCECHNGYQMDPTTGVCKAVGKEPCLIF 443
ObVLDLR CLQNNGGCSHICDNDVIGYECDDTPGLQLIDRKTCDGIDNECLNPGICSOICINLKGKGYKCECHNGYQMDSTTGVCKAVGKEPCLIF 443
SMDLR CLQNNGGCSHICDNDVIGYECDDTPGLQLIDRKTCDGIDNECLNPGICSOICINLKGKGYKCECHNGYQMDSTTGVCKAVGKEPCLIF 443
OmVLDLR CLLNNGGCSHICDNDVIGYECDDTPGLQLIDRKTCDGIDNECLNPGICSOICINLKGKGYKCECHNSYQMDPTTGVCKAVGKEPCLIF 443
CcVLDLR CLLNNGGCSHICDNDVIGYECDDTPGLQLIDRKTCDGIDNECLNPGICSOICINLKGKGYKCECHNGYQMDPTTGVCKAVGKEPCLIF 443
OaVLDLR CLLNNGGCSHICDNDVIGYECDDTPGLQLIDRKTCDGIDNECLNPGICSOICINLKGKGYKCECHNGYQMDPTTGVCKAVGKEPCLIF 445
HnVLDLR RDIRKLTLLERREYTOIVEQLRNTVALDADFQOQIFWADLGGQKAFSMLDKR-DEGSTITKVI-ESVQTPVGI-VAVDWIYKNIYWSDLG 531
ObVLDLR RDIRKLTLLERREYTOIVEQLRNTVALDADFQOQIFWADLGGQKAFSMLDKR-DGSAITKVIDNVHTPVGI-VAVDWIYKNIYWSDLG 531
SMDLR RDIRKLTLLERREYTOIVEQLRNTVALDADFQOQIFWADLGGQKAFSMLDKR-DGSTITKVIDNVHTPVGI-VAVDWIYKNIYWSDLG 531
OmVLDLR RDIRKLTLLERREYTOIVEQLRNTVALDADFQOQIFWADLGGQKAFSMLDKR-DGSHVKKVIDNVHTPVGI-VAVDWIYKNIYWSDLG 531
CcVLDLR RDIRKLTLLERREYTOIVEQLRNTVALDADFQOQIFWADLGGQKAFSMLDKR-ESAHMKVIDNVHTPVGI-VAVDWIYKNIYWSDLG 530
OaVLDLR RDIRKLTLLERREYTOIVEQLRNTVALDADFQOQIFWADLGGQKAFSMLDKR-VGTHKKVIDNVHTPVGI-VAVDWIYKNIYWSDLG 534
HnVLDLR TISVANFNGTKKLVLFSSGLKEPASIAVDPLSGFLYWSDWGEPAKIEKSGMNGVDRQVLVETDIQWPNGITLDLIKSRLYVWDSK 620
ObVLDLR TISVANFNGTKKLVLFSSGLKEPASIAVDPLSGFLYWSDWGEPAKIEKSGMNGIDRQVLVETDIQWPNGITLDLIKSRLYVWDSK 620
SMDLR TISVANFNGTKKLVLFSSGLKEPASIAVDPLSGFLYWSDWGEPAKIEKSGMNGVDRQVLVETDIQWPNGITLDLIKSRLYVWDSK 620
OmVLDLR TISVANFNGTKKLVLFSSGLKEPASIAVDPLSGFLYWSDWGEPATIEKSGMNGVDRQVLVETDIQWPNGITLDLIKSRLYVWDSK 620
CcVLDLR SIAVANFNGTKKLVLFSSGLKEPASIAVDPLSGFLYWSDWGEPAKIEKSGMNGIDRQVLVETDIQWPNGITLDLIKSRLYVWDSK 619
OaVLDLR TISVASFNGTKKLVLFSSGLKEPASIAVDPLSGFLYWSDWGEPAKIEKSGMNGVDRQVLVETDIQWPNGITLDLIKSRLYVWDSK 623
HnVLDLR CSVDLNGDNRRKVLQSPDYLAHPFALTFEDRVFWDGENEA1YGANKFTGSDVITMLASNLNEPQDI1VYHEL IQLSGTNWCDK 709
ObVLDLR CSVDLNGDNRRKVLQSPDYLAHPFALTFEDRVFWDGENEA1YGANKFTGSDVVMLASNLNEPQDI1VYHEL IQLSGTNWCDK 709
SMDLR CSVDLNGDNRRKVLQSPDYLAHPFALTFEDRVFWDGENEA1YGANKFTGSDVVMLASNLNEPQDI1VYHEL IQLSGTNWCDK 709
OmVLDLR CSVDLNGDNRRKVLQSPDYLAHPFALTFEDRVFWDGENEA1YGANKFTGSDVITLASENLNEPQDI1VYHEL IQLSGTNWCDK 709
CcVLDLR CSVDLNGDNRRKVLQSPDYLAHPFALTFEDRVFWDGENEA1YGANKYTGSEVTLASNLNEPQDI1VYHEL IQLSGTNWCDK 708
OaVLDLR CSVDLNGDNRRKVLQSPDYLAHPFALTFEDRVFWDGENEA1YGANKFTGSDVITLASENLNEPQDI1VYHEL IQLSGTNWCDK 712
HnVLDLR GCAFMLCPAPQINKHSPKYTCVCPGGMELAADGFCRPPATISAHPGD-GKTLIHPTHASEGNISTSIHEVNSSAKGSAAWAAILPV 797
ObVLDLR GCAFMLCPAPQINKHSPKYTCVCPGGMELAADGFCRPPATISAHPGD-----EGNVSTSIHEVNSSAKGSAAWAAILPV 777
SMDLR GCAFMLCPAPQINKHSPKYTCVCPGGMELAADGFCRPPATISAHPGD-GKVLVHPTHASEGNVSTSIHEVNSSAKGSAAWAAILPV 797
OmVLDLR GCAFMLCPAPQINKHSPKYTCVCPGGMELAADGFCRPPATISAHPGD-----EANSSTSIHEVNSTARGSTAAWAAILPV 777
CcVLDLR GCAFMLCPAPQINKHSPKYTCVCPGGMELAADGFCRPPATISAHPGD-----EANSSTSIHEVNSTARGSTAAWAAILPV 797
OaVLDLR GCAFMLCPAPQINKHSPKYTCVCPGGMELAADGFCRPPATISAHPGD-----EANSSTSIHEVNSTARGSTAAWAAILPV 779
HnVLDLR AMAAAGGYLMWRNWQLKNNKSMNFDNPVYLKTTTEEDLNIDISRHSASVGHYTPAISVNTTEDDL----- 861
ObVLDLR AMAAAGGYLMWRNWQLKNNKSMNFDNPVYLKTTTEEDLNIDISRHSASVGHYTPAISVNTTEDDL----- 841
SMDLR AMAAAGGYLMWRNWQLKNNKSMNFDNPVYLKTTTEEDLNIDISRHSASVGHYTPAISVNTTEDDL----- 861
OmVLDLR AMAAAGGYLMWRNWQLKNNKSMNFDNPVYLKTTTEEDLNIDISRHSASVGHYTPAISVNTTEDDL----- 847
CcVLDLR AMAAAGGYLMWRNWQLKNNKSMNFDNPVYLKTTTEEDLNIDISRHSASVGHYTPAISVNTTEDDL----- 862
OaVLDLR AMAAAGGYLMWRNWQLKNNKSMNFDNPVYLKTTTEEDLNIDISRHSASVGHYTPAISVNTTEDDL----- 844

```

Figure 4c. Alignment of deduced protein sequences of *Heterotis niloticus* vitellogenin receptor (HnVLDLR) with *Sclerogages formosus* VLDLR (SfVLDLR), *Osteoglossum bicirrhosum* VLDLR (ObVLDLR), *Ancorhynchus mykiss* VLDLR (OmVLDLR), *Oreochromis aureus* VLDLR (OaVLDLR) and *Cyprinus carpio* VLDLR (CcVLDLR). The cysteines conserved across species are in boxes, the highly conserved SDE, EDE, YWS(V)D, FWA(T)D and FDNPVY sequences are underlined, the O-glycosylation domain is double-underlined, the little conserved domains are in gray.

1 M A S S P G S D L P L L Q F Q E V G S S K T G D R S S P G P L S G V Y
1 **ATGGCTAGT**TCCCCAGGAAGTGACCTGCCACTTCTGCAGTTCAGGAAGTGGGCTCCAGCAAGACAGGGGACCGCAGCTCCCCAGGACCCCTCTCGTGTCTAC
36 A G P V P G L A M E S R A V C I P S P Y A D S S H D Y T T L G F Y N P
106 GCTGGTCTGTGCCAGGTCTAGCCATGGAGAGCCGAGCAGTGTGCATCCCTCGCCCTATGCAGATAGCAGCCACACTCTTGGTCTCTATATCCC
71 S M L G Y P G S V P D S P S V R P P L S P A I Y W S P Q S H P S Q L P
211 TCCATGCTGGGGTACCCGGGCTCTGCCCTGACAGCCCTCTGTGCGGCCACCCCTAAGCCCTGCTATCTACTGGTCACCACAGAGCCATCCTAGCCAGTGGCC
106 P L T L H C Q Q P L M Y G E P T R T P W V E P K A Q D H N L V E S S K
316 CCGCTCACCTTGCATTTGTCAGCAGCCCTGTATGTATGGTGGAGCCACTCGGACCCCATGGTGGAGCCCAAGCCAGGACCACAACCTTGGTGGAGAGCAGTAA
141 L A G R R P L E G D E A L S S S A A C L T A K G D M H F C A V C H D Y
421 CTGGCGGGGCGACGGCCACTAGAAGGAGATGAGGCACTCAGCTCTTCTGCAGCCTGCTTAACGGCAAGGGCGATATGCATTTCTGCGCTGTCCACGACTAT
176 A S G Y H Y G V W S C E G C K A F F K R S I Q G H N D Y I C P A T N Q
526 GCCTCAGTTTACTATGGTGTCTGGTCTTTCGAGGGGCTGCAAGCCCTCTTTAAGAGGAGCATTCAAGGGCACAATGACTACATTTGCCAGCGACCAACCCAG
211 C T I D K N R R R K S C Q A C R L R K C Y E V G M M K C G V R R E R C S
631 TGCACCATCGACAAGAACCCTCGCAAGAGCTGCCAGGCTGCCGCTCGCAAGTGTCTACGAAGTGGGCATGATGAAGTGGGATGAGACGTGAGCGTGCAGC
246 Y R G V R H R R V P Q I R E V M L G S G S R T Q R R L E S S L P P T K
736 TACCCTGGGGTGCACACGGGCTGTACACAGATCCGAGAAGTGTGGGGTTCAGGCTCTAGGACCCAGAGGCGACTGGAGAGCAGCTCCCCCAACGAAG
281 S F Q S L A L T P E Q L V L R I I E A E P P E I Y L M K D M K K P F T
841 AGTTTCCAGTCCCTGGCGCTGACCCTGAACAGCTGGTGTGGCCATCATTGAGGCAGAACCCCAAGATCTACTTGTGAAGGACATGAAGAAGCCATTTACC
316 E S S M M M S L T N L A D K E L V L M I S W A K K I P G F V E L N L S
946 GAGACAGCATGATGATGTCTTACTAACCCTGGCAGACAGGAGTGTGGTCCCTGATCAGTGGGCCAAAAGATCCAGGGTTTGTGGAGCTCACCTGTGCA
351 D Q V H L L E C C W L E V L M L G L M W R S V D H P G K L I F S P D L
1051 GATCAAGTGCACTGTGGAGTGTGGTGGTGGAGGTGTGATGTGGGCTGATGTGGAGGTGTGTGACCATCCGGGAAGCTTATCTCTCCCTGACCTC
386 K L N R D E G S C V E G I M E I F D M L L A A T A R F R E L K L Q R E
1156 AAGCTTAACAGGGATGAGGGCAGCTGTGTGGAGGGATCATGGAGATCTTTGACATGTTGCTGGCAGCTACCGCCGATTCGGGAAGTGAAGCTGCAGAGGGAA
421 E Y V C L K A L I L L N S N M C V N S P E T P E E L E S R A K L L R L
1261 GAGTATGCTGTCTCAAGGCGCTGATCCTGCTCAACTCCAACATGTGTGAACTCTCCAGAGACACCTGAGAGCTGGAGAGTGGGCCAAGCTGTGCGGCTG
456 L D A V T D A L V W A I S R K G L T F Q Q Q S A R L A H L L M L L S H
1366 CTGGACGCCGTCACAGACGCACTGGTGTGGGCCATCTCCAGGAAGGGCCTCACCTTTCAGCAACAGTCAGCTCGCCTCGCTCATCTACTGATGCTGTGTACAC
491 I R H L S N K G M E H L S N M K M K N V V P L Y D L L L E M L D A N T
1471 ATCCGCCATCTCAGTAACAAGGAAATGGAGCACCTCTCCAACATGAAGATGAAGAATGCTGCGCGCTGTACGATCTGCTCCTGGAGATGCTGGACGCCAACAC
526 M H S S R V H Q R P P P G D I N P Q E S S S H H T A E N K P L Q Q *
1576 ATGCACAGCTCCCTGTGCATCAGCGCCCCACCAGGGGACATCAATCCCCAGGAGTCTCTGTCACACCAACACAGCAAAAACAGCCCTGCAGCAGTAG

Figure 5a. Nucleotide sequence and deduced protein sequence of *Heterotis niloticus* (Hn) estrogen receptors beta (ER β _X1) identified in the liver and ovary of *Heterotis niloticus*. The start codon and the stop codon are in bold type. The A/B region in N-terminal position is followed by the DNA-binding domain (DBD) located in the boxes and containing a double-underlined P-box and a single-underlined D-box. The ligand-binding domain (LBD) is underlined and followed by an F-domain in C-terminal position. The amino acids implied in the binding with estradiol are circled.

(ERβ_X2)

```
1 M A S S P G S D L P L L Q F Q E V G S S K T G D R S S P G P L S G V Y
1 ATGGCATAGTTCCCCAGGAAGTACCTGCCACTTCTGCAGTTCAGGAAGTGGGCTCCAGCAAGACAGGGGACCGCAGCTCCCCAGGACCCCTCTCGGTGTCTAC
36 A G P V P G L A M E S R A V C I P S P Y A D S S H D Y T T L G F Y N P
106 GCTGGTCTGTGCCAGGTCTAGCCATGGAGAGCCGAGCAGTGTGCATCCCCTCGCCCTATGCAGATAGCAGCCAGACTACACCCTCTGGCTTCTATATATCCC
71 S M L G Y P G S V P D S P S V R P P L S P A I Y W S P Q S H P S Q L P
211 TCATGCTGGGGTACCGGGCTGTACCCAGACCCCTGTGCGGCCACCCCTAAGCCCTGCTATCTACTGGTACCACAGAGCCATCCTAGCCAGCTGGCC
106 P L T L H C Q Q P L M Y G E P T R T P W V E P K A Q D H N L V E S S K
316 CCGCTCACCTTGCATGTGACGAGCCCTGTATGTATGGTGGAGCCACTCGGACCCATGGGTGGAGCCCAAGGCCAGGACCAACTTGGTTGAGAGCAGTAA
141 L A G R R P L E G D E A L S S S A A C L T A K G D M H F C A V C H D Y
421 CTGGCGGGCGCAGGCCACTAGAAGGAGATGAGGCACCTAGCTCTCTGCAGCCTGCTTAACGGCAAGGGCGATATGCATTTCTGGCTGTCTGCCAGCAT
176 A S G Y H Y G V W S C E G C K A F F K R S I Q G H N D Y I C P A T N Q
526 GCCTCAGGTTACTACTATGGTGTCTGGTCTGCGAGGGCTGCAAGGCCCTCTTTAAGAGGAGCATTCAAGGGCAACATGACTACATTTGCCAGCGACCAACCAG
211 C T I D K N R R K S C Q A C R L R K C Y E V G M M K C G V R R E R C S
631 TGCACCTCGACAAGACCTCGCAAGAGCTGCCAGCCCTGCGAAAGTGTACGAAGTGGGCATGATGAAGTGGAGTGGAGCGTGGAGCGCTGCAGC
246 Y R G V R H R R V P Q I R E V M L G S G S R T Q R R L E S S L P P T K
736 TACCCTGGGGTCCGACCCGGCTGTACCCAGCATCCGAGAAGTGTATGGGGTCCAGGCTCTAGGACCCAGGCGACTGGAGAGCAGCCCTCCCCCAACGTA
281 S F L A L T P E Q L V L R I E A E P P E I Y L M K D M K K P F T
841 AGTTTCCAGTCCCTGGCGTACCCTGAACAGCTGGTGTGGCCTCATTTGAGGAGACACCCAGAGATCTACTTGTGAAGGACATGAAGAAGCCATTTACC
316 E S M M M S L T N L A D K (E) L V L M I S W A K K I P G F V E L N L S
946 GAGACGCATGATGTCTTACTAACCCTGGCAGCAAGAGTGGTCCCTATGATCAGCTGGGCCAAAAGATCCAGGGTGTGGAGCTCACACCTGTCA
351 D Q V H L L E C C W L E V L M L G L M W (R) S V D H P G K L I F S P D L
1051 GATCAAGTGACCTGCTGGAGTGTGTGGTGGAGGTTGTATGCTGGCCCTGATGTGGAGTCTGTGACCATCTCCGGGAGGTTATCTCTCCCTGACCTC
386 K L N R D E G S C V E G I M E I F D M L L A A T A R F R E L K L Q R E
1156 AAGCTTAACAGGATGAGGGCAGCTGTGGAGGGATCATGGAGATCTTTGACATGTTGCTGGCAGCTACCCGCCGATCCCGGAATCCCGGAATGAAGTGCAGAGGGAA
421 E Y V C L K A L I L L N S K T P E E L E S R A K L L R L L D A V T D A
1261 GAGTATGCTGTCTCAAGCGCTGTACCTCACTCAAGACACCTGAGGAGCTGGAGAGTGGGCCAAGCTGCTGGCGTGTGGAGCGTCAACAGCAGCA
456 L V W A I S R K G L T F Q Q Q S A R L A H L L M L L S H I R H L S N K
1366 CTGGTGGGCTCTCCAGGAAGGCCCTCACCTTTCAGCAACAGCTCAGCTCGCCTCGCTCATCTACTGATGCTGCTCACACATCCGCCATCTCAGTAACAAA
491 (M) (E) (H) (S) (N) (M) (V) (P) (L) (Y) (D) (M) (L) (L) (E) (M) (L) (D) (A) (N) (M) (H) (S) (R) (V)
1471 GGAATGGAGCACCTCCACATGAAGATGAAGAATGTCGTGCCGCTGTACGATCTGCTCCTGGAGATGCTGGAGCAACACCATGCACAGCTCCCGTGTGCAT
526 Q R P P P G D I N P Q E S S S H H T A E N K P L Q Q *
1576 CAGCGCCCCCACCAGGGACATCAATCCCCAGGAGTCCCTCGTACACACACAGCAGAAAACAAGCCCTGCAGCAGTAG
```

(ERβ_X3)

```
1 M H A S L C P S K L A G R R P L E G D E A L S S S A A C L T A K G D M
1 ATGCATAGTTCCTCTGTCCCAGTAACTGGCGGGCGACGGCCACTAGAAGGAGATGAGGCACCTCAGCTCTTCTGCAGCCTGCTTAACGGCAAGGGCAGTATG
36 H F C A V C H D Y A S G Y H Y G V W S C E G C K A F F K R S I Q G H N
106 CATTCTCGCTGTCTGCCAGCTATCCCTCAGGTTACTACTATGGTGTCTGGTCTTGGAGGGTCTCAAGGCCCTCTTTAAGAGGAGCATTCAAGGGCACAAT
71 D Y I C P A T N Q C T K M I D K N R L R K S C Q A C R L R K C Y E V G M M K
211 GACTACATTTGCCCGGCAACACCAAGTGCACCATCGACAAGACCCGTCGAAGAGTGCAGGCCCTGCCCTCGCAAGTGTACGAAGTGGGCATGATGAAG
106 C G V R R E R C S Y R G V R H R R V P Q I R E V M L G S G S R T Q R R
316 TCGCGAGTGGAGCTGAGCCCTGACCTACCCGTGGGGTGGCAGACCCGGCTGTACCCAGATCCGAGAAGTGTATGGGGTCCAGGCTCTAGGACCCAGAGCGCA
141 L E S S L P P T K S F Q S L A L T P E Q L V L R I E A E P P E I Y L
421 CTGGAGACAGCCTCCCCCAACGAAGAGTTTCAGTCCCTGGCGTACCCTGAACAGCTGGTGTGGCCTCATTTGAGGAGACACCCAGAGATCTACTTG
176 M K D M K K P F T E S S M M M S L T N L A D K E L V L M I S W A K K I
526 ATGAAGACATGAAGAAGCCATTTACCGAGAGCAGCATGATGATGCTCTTACTAACCCTGGCAGCAAGGAGTGGTCCCTATGATCAGCTGGGCCAAAAGATC
211 P G F V E L N L S D Q V H L L E C C W L E V L M L G L M W R S V D H P
631 CCAGGTTTGTGGAGCTCAACCTGTCAGATCAAGTGCACCTGCTGGAGTGTGGTGGTGGAGGTTGATGCTGGCGCTGATGCTGGAGGTTGTTGACCATCTC
246 G K L I F S P D L K L N R D E G S C V E G I M E I F D M L L A A T A R
736 GGAAGCTTATCTTCCCTGACCTCAAGCTTAACAGGATGAGGGCAGCTGTGGAGGGATCATGGAGATCTTTGACATGTTGCTGGCAGTACCGCCGA
281 F R E L K L Q R E E Y V C L K A L I L L N S N M C V N S P E T P E E L
841 TTCCGCAACTGAAGCTCGAGAGGGAAGATGATGCTGTCTCAAGCGCTGATCCTGCTCAACTCCAACATGTGTGAAGTCTCCAGAGACACCTGAGGAGCTG
316 E S R A K L L R L L L D A V T D A L V W A I S R K G L T F Q Q Q S A R L
946 GAGAGTGGGCCAAGCTGCTGGGCTGTGGAGCGCTCACAGACGACTGGTGGGCCATCTCCAGGAAGGGCTCACCTTTCAGCAACAGCTCAGCTCGCCTC
351 A H L L M L M L L S H I R H L S N K G M E H L S N M K M K N V V P L Y D L
1051 GCTCATCTGATGCTGCTCACACATCCGCCCTCAGTAAACAAGGAATGAGGACCTTCCAACATGAAGATGAGATGCTGCTGGCTGCTGCTGCTGCTG
386 L L E M L D A N T M H S S R V H Q R P P G D I N P Q E S S S H H T A
1156 CTCTGGAGATGCTGGAGCAACACCATGCACAGCTCCCGTGTGCATCAGCGCCCCCACCAGGGGACATCAATCCCCAGGAGTCCCTCGTACACACACAGCA
421 E N K P L Q Q *
1261 GAAAACAAGCCCTGCAGCAGTAG
```

Figure 5b. Nucleotide sequences and deduced protein sequences of *Heterotis niloticus* (Hn) estrogen receptors beta (ERβ_X2 and ERβ_X3) identified in the liver of *Heterotis niloticus*. The start codon and the stop codon are in bold type. The A/B region in N-terminal position is followed by the DNA-binding domain (DBD) located in the boxes and containing a double-underlined P-box and a single-underlined D-box. The ligand-binding domain (LBD) is underlined and followed by an F-domain in C-terminal position. The amino acids implied in the binding with estradiol are circled.

(ER β _X4)

1 M A S S P G S D L P L L Q F Q E V G S S K T G D R S S P G P L S G V Y
1 **ATGGCTAGTTC**CCCCAGGAAGTGCACCTGCCACTTCTGCAGTTCAGGAAGTGGGCTCCAGCAAGACAGGGGACCGCAGCTCCCCAGGACCCCTCTCGTGTCTAC
36 A G P V P G L A M E S R A V C I P S P Y A D S S H D Y T T L G F Y N P
106 GCTGGTCTGTGCCAGGCTTAGCCATGGAGAGCCGAGCAGTGCATCCCTCGCCCTATGCAGATAGCAGCCAGCAGTACACCACCTTTGGCTTCTATAATCCC
71 S M L G Y P G S V P D S P S V R P P L S P A I Y W S P Q S H P S Q L P
211 TCCATGCTGGGGTACCCGGGCTGTGCCTGACAGCCCTCTGTGGGCCACCCCTAAGCCCTGTATCTACTGGTCCACCAGAGCCATCCTAGCCAGCTGCC
106 P L T L H C Q Q P L M Y G E P T R T P W V E P K A Q D H N L V E S S K
316 CCCTCACCCTTGCATGTGCAGCAGCCCTGTATGTATGGTGGAGCCACTCGGACCCCATGGTGGAGCCCAAGGCCAGGACCACACTGGTTGAGAGCAGTAA
141 L A G R R P L E G D E A L S S S A A C L T A K G D M H F F C A V C H D Y
421 CTGGCGGGCAGCCGCTAGAAGGAGATGAGGCACTCAGCTCTTCTGCAGCTGCTTAACGGCAAAGGGCGATATGCATTTCTGGCGTGTGCCAGCAGTAT
176 A S G Y H Y G V W S C E G C K A F F K R S I Q G H N D Y I C P A T N Q
526 GCCTCAGGTTATCACTATGGTGTGGTCTTGGCAGGGTCAAGGCCCTTTAAGAGGAGCATTCAAGGCCAATGACTACATTTGCCAGGCCCAACCAG
211 C T I D K N R R K S C Q A C R L R K C Y E V G M M K C G V R R E R C S
631 TGACCATCGACAGAACCGTCCGAAGAGCTGCCAGCTGCCCGCTCGCAAGTGTACGAAGTGGGCATGATGAAGTGGCGGATGAGACGTGAGCAGCTGCAGC
246 Y R G V R H R R V P Q I R E V M L G S G S R T Q R R L E S S L P P T K
736 TACCCTGGGGTGCACACCCGGCTGTACCACAGATCCGAGAAGTGTATGGTGGGTGAGGCTTAGGACCCAGGGCGACTGGAGAGCAGCTCCCCCAACGAAG
281 S F Q S L A L T P E Q L V L R I E A E P P E I Y L M K D M K K P F T
841 AGTTTCCAGTCCCTGGCGCTGACCCCTGAACAGCTGGTGTGGCCATCATTGAGGCGAAGCCCAAGAGATCTACTGTAGGAGACATGAAGGCCATTTACC
316 E S S M M M S L T N L A D K E L V L M I S W A K K I P G F V E L N L S
946 GAGAGCAGCATGATGATCTCTTACTAACCCTGGCAGACAAGGAGTGGTCCCTCATGATCAGCTGGGCCAAAAGATCCCAGGGTTTGGAGCTCAACCTGTCA
351 D Q V H L L E C C W L E V L M L G L M W R S V D H P G K L I F S P D L
1051 GATCAAGTGCACCTGCTGGAGTGTGCTGGCTGGAGGTGTGATGCTGGCCCTGATGTGGAGTCTGTGACCATCCTGGGAAGCTTATCTTCTCCCTGACCTC
386 K L N R D E G G S C V E G I M E I F D M L L A A T A R F R E L K L Q R E
1156 AAGCTTAAACAGGATGAGGGCAGCTGTGTGGAGGATCATGGAGATCTTTGACATGTTGCTGGCAGCTACCCCGCATTCGGGAAGTGAAGCTGCAGAGGAA
421 E Y V C L K A L I L L N S N M C V N S P E T P E E L E S R A K L L R L
1261 GAGTATGCTGTCAAGGCGCTGATCCTGCTCAACTCCAACATGGTGTGAAGTCTCCAGAGACACCTGAGGAGTGGAGAGTGGGCCAAAGCTGTGGCGGTG
456 L D A V T D A L V W A I S R K G L T P F Q S A R L A H L L M L L S H
1366 CTGGAGCCGCTCACAGAGCAGTGGTGGGCTCATCTCAGGAAGGGCTCACCTTCAGCAACAGTCACTGCCTCGCTCATCTACTGATGCTGCTGCACAC
491 I R H L R V P V V W E F P V F P G H *
1471 ATCCGCCATCTCAGGGTTCCGGTGGTGGGAATTTCTGTATTCCCTGGGCAT**TAA**

(ER γ)

1 M L L S L V Y S S A T P I N I P S P L M D N C Q D Y S V Q Q Y A C P G
1 **ATGTTGCTG**CCCTGGTCTACAGCTCTGCTACGCCATCAACATCCCATCTCCATTAATGGACAACCTGCCAAGATTAAGTACAGCAATATGCTTCCCTGGT
36 L S D S S F L D P P F D W K P D S N F L S I V K K I T P P S S P L A E
106 CTCAGTGATCCAGCTTTTTGGACCCACCGTTTGGTGAACCCGACAGCAATTTCTTGTCAATAGTCAAGAAAATACACCACCCAGCAGCTCCTGGGTGAG
71 P E P Q S L F S T D K Q Q R R K R K T N G M C R V K S L Q P F P G D
211 CCTGAGCCCAAGTCTTTCTCTACTGATAAACAGCAGCAGAGAAGAAGGAAACGAATGGCATGTGCAGAGTGAATCACTGCAGCCTTTCCAGGGGAT
106 E T E K R L C F V C K D Y A S G Y H Y S V W S C E G C K A F F K R S I
316 GAGACAGAAAACGCTTGTGTTTTGTGTCAAGGACTATGCTCAGGATACCACTACAGTGTGGTCCCTGTGAGGGTCAAGGCCCTTTTTAAGAGGAGCATC
141 Q G P T D Y I C P A N N Q C T I D K S R R K S C Q A C R L R K C Y E V
421 CAAGGGCTACTGATATATCTGTCCAGCAACACCAAGTGCACCATGTGACAAGAGCCCGCCAAAGCTGCCAGGCTTGTGCTTTCGCAAGTGTATGAAGTG
176 G M V K N G V R R P S Q E V Q G I R M C Q H L S M G E A S L
526 GGCATGGTCAAGAATGGTGTCCGAAGAACGCTGCTCCCTCCAGGAGTGCAGGATACGATGTGCCAGCAGCTTTCAGTAATGGGGGAAGAGCCAGTCTT
211 L S Q E K S H L S A S L E E T Q Q L L F S P K Q L I S R F L E I S V E P P
631 CTCTCACAAGAGAAGTCCACCTTTCTGCATCTTTGGAGGAGACCAACAAGTGTGTTCCAGCCAAAGCAGCTCATTTCTCGCTTTTGGAGGTAGAACCTCCA
246 D I Y L M W D L K K P F T E V S M M T S L T H L A D K E L F Y M I S W
736 GACATTTACTGTGAGGAGCTGAAAACCCCTTACAGAGGTACAGATGATGACATCACTCACTCAGTGGCAGCAAAAGAGCTTTTTACATGATCAGCTGG
281 A K K I P G F T E L G L S D Q V H L L D C C W L E V L M L G L I W R S
841 GCTAAAAGATCCCGGCTTCACTGAGCTCGGCCATCAGACCAGGTACACTACTAGACTGTGCTGGCTGGAGGTGCTAATGGTGGGCTGTGCTGGAGGCTT
316 V D H P G K L I F S P D L I L S R D E G S C I E G I L E I F D M L L A
946 GTGGACCACCTGGAAAGCTGATCTTCTCCCTGACCTATCCTTAGCAGGGATGAGGGCAGCTGCATTTGAGGGGATTTGAGAGATCTTTGACATGCTGCTTGCA
351 A T S R F R E L R L Q K E E Y V C L K A L V L L N S S I Y L Y T P N G
1051 GCAACTTCTCGTTCGGGAGCTGCGGGTACAAAAGGAGGAATATGTCTGCCTGAAGCCCTGGTCTCCTCAACTTACGATATACCTTATACCCCAATGGT
386 G Y E H Q S R A K L Q Q L L N A V T D A L I W V I A K L G L S F Q Q Q
1156 GGGTATGAGCATCAGAGCAGGGCCAAAGCTACAACAGCTGTGAATGCTGTGACTGATGACCTTATATGGTATCGCTAAGCTGGGACTGAGCTTCCAGCAGCAG
421 S A R L A H L I M L L S H I R H V S N K G M D H L H C M K M K K V V P
1261 TCCGCCCAGTGGCCATCTCATCTGTTGTGCACATCCGCCATGTGATCAACAAAGGAATGGACCACCTGCAGCTGCATGAAGTGAAGAAAGTAGTGCCA
456 F Y D L L L E M L D A H V M H N S R I L S P E P S G *
1366 TTCTATGATCTCTTGGAAAGTGGAGCCCATGTTATGCACAAAGCCCATTTCTCAGTCCAGCTCAGAG**TGA**

Figure 5c. Nucleotide sequences and deduced protein sequences of *Heterotis niloticus* (Hn) estrogen receptors beta (ER β _X4) and gamma (ER γ) identified in the ovary of *Heterotis niloticus*. The start codon and the stop codon are in bold type. The A/B region in N-terminal position is followed by the DNA-binding domain (DBD) located in the boxes and containing a double-underlined P-box and a single-underlined D-box. The ligand-binding domain (LBD) is underlined and followed by an F-domain in C-terminal position. The amino acids implied in the binding with estradiol are circled.

(ERR α)

1 M S S R D R R F N I Y I K A E P S S P E G G R D G R T S P G G A S S D
1 **ATG**TCTTCCCGGATCGACGCTTCAATATTTACATTAAGGCTGAGCCAGCAGTCCAGAGGAGGAAGAGATGGGAGAACCAGCCCTGGAGGGGCATCTCCGAT
36 S S S H S A V G V N G G R S T E R Y S P P L C T P T L H C P F K E E G D
106 TCGTCCACAGTGTGTGGGGTAAATGGAGGCAGGACAGAGCGGTACTCCCTCCCTCTGCACACCAACCTGCATGGCCCTTTAAAGAGGAGGGCGAT
71 G E E G S A G S G G G R C K Y A L S T L P K R L C L V C G D V A S G Y
211 GGAGAGGGGATCTCGAGGACGGGAGCGGAGGTGCAAGTACGCCACTCCCAAGCGCTGTGTCTCGTGTGGGATGTGGCATCAGGCTAC
106 H Y G V A S C E A C K A F F K R T I Q G N I E Y S C P A S N E C E I T
316 CACTACGGCGTGGCCCTCTGTGAGGCTTCTTCAAGAGGACATTCAGGGAACTTGAATACAGCTGCCAGCATCCACAGTGGCAATCACC
141 K R R R K A C Q A C R F T K C L K V G M L K E G V R L D R V R G G R Q
421 AAGCGCCAGGAAGGCTGCGAAGCTTGTTCGCTTACCAGTGCCTCAAAGTGGGAATGCTGAAAGAGGGTGTTCGCTGGATCGAGTGGAGGAGGAAGACAG
176 K Y K R R P E V E S T T Y Q S P T N Q P S G K E V D K G P S N V I V S
526 AAGTACAAAAGCGCCAGAGTGGAGAGCACCACATACCAAAGTCTACGAATCAGCCCTCAGGAAAGGAAGTGTGATAAAGTCCCTCAATGTCTTCTCC
211 H L L V A E P E K L F A M P D P L Q S D T T L R T L T T L C D L A D R
631 CATCTCGTGGTGTGAGCCAGAGAAGTGTGGCCATGCCAGACCCCTCAGTGCAGACACCACCTCGCCACCTCACCACCTTTGTGACTTGGCTGATCGC
246 E L V V I I G W A K H I P G F L S L T L S L A D Q M S V L Q S V W L E V L
736 GAGCTGTGTGCATATTTGGCTGGCTAAACACATTCGGGCTTCTGTCTTCTGTCTGGCGGACAGATGTGAGTGTTCGATGCGGTGGAGTGGAGTCTC
281 V L G V A F R S L S C E D E V V F T A E D F V L D E E L S R I A G L G E
841 GTACTCGTGTTCGCTTCCGCTCTGAGCTCAGGAGGAGGTGGTGTTCGCCAGGACTTTGTCTGGATGAGGAGCTTCCCGTATTTGCTGGCTGGCGGAG
316 L S A A I S Q L A R R Y R A C L Q L D Q E E F V M L K A I A L T L N S D S
946 TTAAGTGCAGCATTAGCAACTGGCCCGCAGTACCGTGCCTTCAACTAGACAGGAGGAGTGTTCATGTTGAAGGCTATTGGCATCACAACTCATGATCT
351 V Y I E D M E A V Q T R Q K L R F D L L H L E L E S Q R R P E D P R A
1051 GTGTACATTGAGGACATGGAGGCTGTGCAGAAGCTACGAGACCTCTACACCAGGCCCTGTAGAGCTGGAAGGCCAGCGTAGGCCAGAAAGCCCGGAGGGCA
386 G R L L L T L P L L R Q T A S R A L S T F Y N I K T R G G V P M H K L
1156 GGCCGCTCCTGCTACCTCGCCCTGTCAAGCAGTGCACGAGTGCACGCGTCCCTTTCCACCTTCTACAACATCAAGACCCGTTGGTGGTGTGCCATGCACAACTC
421 F L E M L E A M M D S P *
1261 TTTCTCGAGATGCTGGAGCCATGATGGACTCGCCG**TAA**

(ERR γ _X1)

1 M A A D D R H L P S S C G S Y I K T E P S S P S S V I D T V S H H S P
1 **ATG**GCTGCAGATGACCCGACCTCCCTCCAGTGTGGTCTTACATCAAGACGGAGCCATCGAGCCCTCATCGGTCACTCGACAGTGCAGCCACCATAGCCCC
36 S G N S D A S G G Y V S A M N S H S N G L D S P P M F T P A C S G L C G G G
106 AGTGGCACTCGAGCCAGCGGTGCTAGCTCAGCGCATGAACAGCCACTCCAATGGCTGGACTCGCCACCCATGTCACGCCAGCGGCTGGAGCGGG
71 A C R K R Y D D G S T I M E D S P I K C E Y M L N S I P K R L C L V
211 GCCTCGCCAGCGCTATGATGACTGCTCCAGCACCATTATGGAGACTCGCCCATTAAGTGGCAATATATGTTGAATCCACCCAGAGGCTGTGCCTAGTC
106 C G D I A S G Y H Y G V A S C E A C K A F F K R T I Q G N I E Y S C P
316 GTGGAGACATTCGCTCAGGATACCACTACGAGTGGCTTCTGAGGCTGCBAAGCCCTTTTAAAGGACAAATCAAGACCCATCGAGTGCAGCTGCTCC
141 A T N E C E I T K R R R K S C Q A C R F M K C L K V G M L K E G V R L
421 CTTACAACAGGAGTGTGAGTCAACAGCAAGCGCAAGCTTCCGCTGCTCATGAAGTGTCTCAAAGTGGGATGTGAAAGGAGTGTGCGTCTG
176 D R V R G R Q K Y K R R M D A E N N A Y L G L T L P P P A K K P L T
526 GACCGTGTGCTGGCGCCGAGAAAGTACAAGAGGAGGATGGATGCCAAAACAACGCCATCTTGGCCCTCACACTACCGCCCTGCCAAAAGCCATGACA
211 K I V S H L L V A E P E K I Y A M P D P T M P E S D I K A L T T L C D
631 AAGATTGTGTCACCTTGTGGTGGCAGAGCCAGAGAAGATCTACGCCATGCCAGACCCACCATGAGAGTGCATCAAGCCCTAAGCCACTTTGTGGCAG
246 L A D R E L V V I I G W A K H I P G F S S L S L G D Q M S L L Q S A W
736 TTTGCTGACCGAGAGTGGTGTTCATCGCTGGGCAAGCATCCAGGCTTCTCCAGTCTGTCTTTAGGAGACAGATGAGTCTACTCGAGCGCTTGG
281 M E I L I L R V Y R S L P F E D K L V Y A E D Y I M D E Y I S K L A
841 ATGGAGATCTGTATCTTGGGGTGTGTACCGATCATGGCCCTCGAGGACAAGTGTGTATGCTGAGGACTACATCATGAGCAGGAGCAGTGCAACTCGCC
316 G L L D L N N A I L Q L V K K Y K S M K L E K E S L S P S K Q L R S Q
946 GGTCTGTGGAACCTCAACAATGCCATCCTCGAGTGGTGAAGAAATCAAGAGCATGAAGTGGAGAAGGAGAGTGTGTCAACCTCAAAGCAATTGCGCTCGCAA
351 I Q T P C T *
1051 ATTACAGATCCATGCACAT**AG**

(ERR γ _X2)

1 M D S V E L S L P E F F S F H S E Q E L L C R M S S K E R R I E S S C
1 **ATG**GATCTGTAGAACTTCCCTGCCTGAGTCTTCTCCTTCACTCTGAACAAGAGTCTCTGCAGGATGTCAAGCAAGGAGCGCCGATTGAGTCCAGTGGC
36 P S Y I K T E P S S P A S L A D S V N H H S P G G S S D A S G S Y S S
106 CCCTCATACATCAAGACAGAACCCTGCTCCCTCCTCGTGCAGCGTGAACCCACAGCCCGGTTGGTCTTCAGACGCCAGTGGCAGTACAGCTCA
71 A T N G H P N G L D S P T L Y G Q A G S L G P N V A G A T C A G E D C
211 GCCACAACGGGACCCCAAGCGCTAGACTCCCCAGCGTGTATGTTGAGCAGGAGCTAGGCCCAATGTGGCCGGGCGCCACCAAGCGTACGAGGACTGC
106 S S T M A E D S Q I K C E Y M L N S M P K R L C L V C G D V A S G Y H
316 CTPAGCAATGGCTGAGACTCACAGATCAAGTGGGAGTACATGCTCAACTGCCCCAAGCGTCTGTGCTGGTGTGGCGATGTGGCATGTGGCTTACCAC
141 Y G V A S C E A C K A F F K R T I Q G N I E Y S C P A T N E C E I T K
421 TACGGAGTGGCTTGTGAAGCTGCAAGGCTTCTTCAAGCGGACATCAAGGAACATCGAGTACAGCTGCGCCGCCACCAACGAGTGTGAGATCACTAAG
176 R R R K S C Q A C R F M K C L T V G M L R E G V R L D R V R G G R Q K
526 AGGAGCGGAAGTCTTGGCCGCTGCGCTTATGAAAGTGTCTACTGTGGAAATGCTGCGGGAAGTGTCCGCTCGACAGATGCGTGGCCGCGGAGAG
211 Y K R E I D A E N S P Y L N P Q L A L P P K K P Y N K V V S H L L V A
631 TACAAGCGCCGATAGACGCGGAGAACGCCCTACCTGAACCCACAGCTGGCCCTGCCACCTAAAAAACATACAACAAAGTGGTGTCCACCTCCTGGTGGCA
246 E P E K I Y A M P D P T V P D S D I K A L T T L C D L A D R E L V V N
736 GAACCTGAGAAATCTATGCACTGCTTCCAGTCCCGACAGGAGATCAAGGCGCTGACCACACTCTCGGACCTGGCTGACCGGAGCTGGTGGTCAAC
281 I G W A K H I P G R S L P S T L E K A V D P S P W P H L V N Y G G A P
841 ATCGGCTGGGCAACACATCCAGTGGAGGCTTCCAGTACCTTGAGAAAGCAGTGGACCCCTCACCTTGGCCACATCTGTTAAACTATGGAGGAGCTCA
316 N R H W S N S G I L V A L L C Y R I L V L V L F L F L C I S L H V
946 AACAGACTGGTCTAACAGTGGCATAGTAAAGCTGGTGTGGCTTGTGTCTACAGGATCTCGTCTTATTTCTTTTAACTCTGCATTTCTTTACATGTC
351 R I F C R V C L Y I V N L G A V S *
1051 AGAATTTTTCGAGATGCTGCCTTATATAGTTAACTTGGAGCAGTAA**AGT****AG**

Figure 6. Nucleotide sequences and deduced protein sequences of the receptors related to *Heterotis niloticus* estrogen receptors alpha (HnERR α) and gamma (ERR γ) (HnERR γ _X1 and HnERR γ _X2). The start codon and the stop codon are in bold type. The A/B region in N-terminal position is followed by the DNA-binding domain (DBD) located in the boxes and containing a double-underlined P-box and a single-underlined D-box. The ligand-binding domain (LBD) is underlined and followed by an F-domain in C-terminal position.

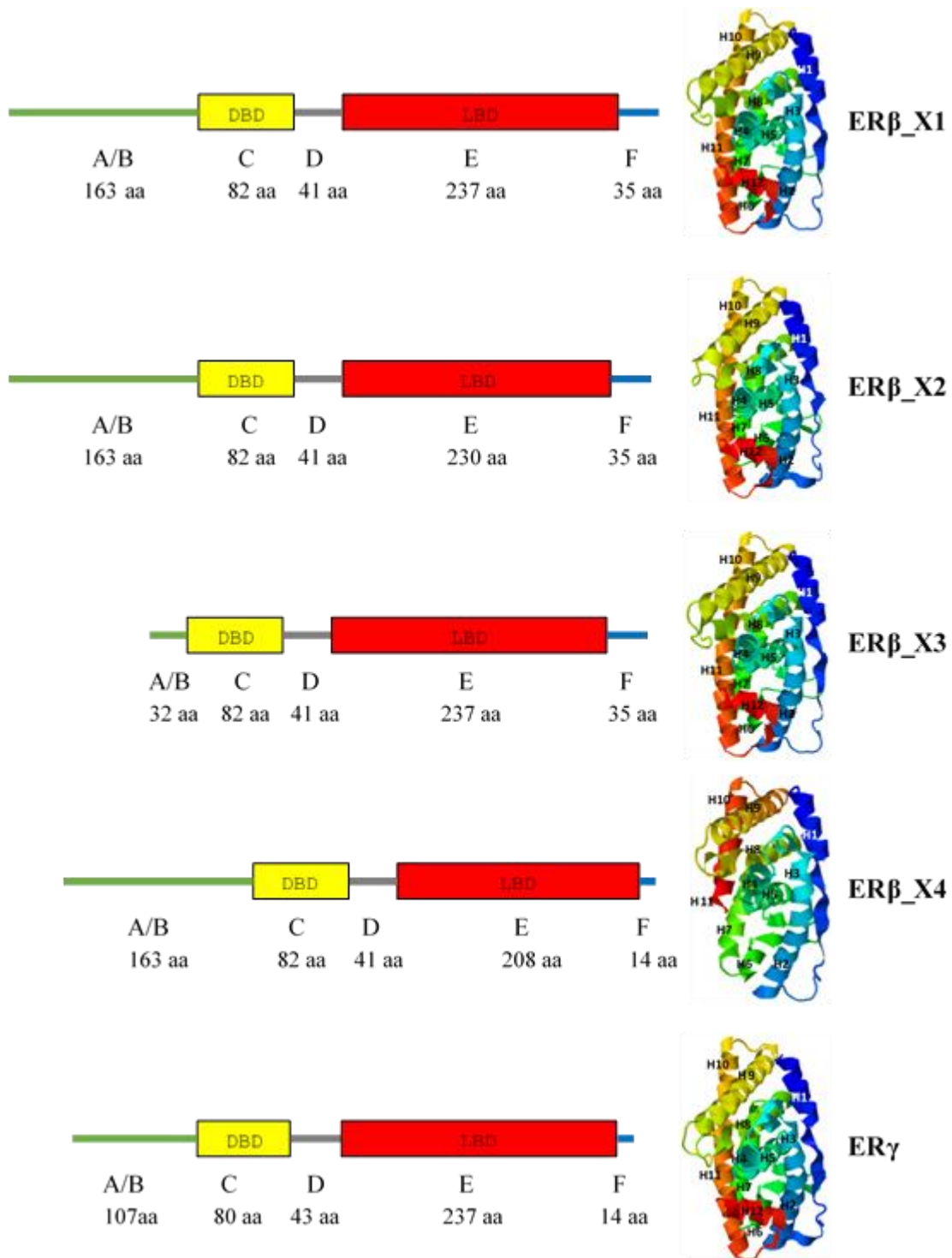


Figure 7. Secondary structure of *Heterotis niloticus* estrogen receptors beta (ERβ_X1, ERβ_X2, ERβ_X3 and ERβ_X4) and gamma (ERγ), with the five characteristic domains and the corresponding numbers of amino acids (aa). The α -helix structure of the LBD obtained with I-TASSER (<https://zhanglab.cmb.med.umich.edu/I-TASSER/>) shows that the 12 helices (H1 to H12) are conserved in ERβ_X1, ERβ_X2, ERβ_X3 and ERγ, while H11 is reduced and H12 is absent in ERβ_X4.

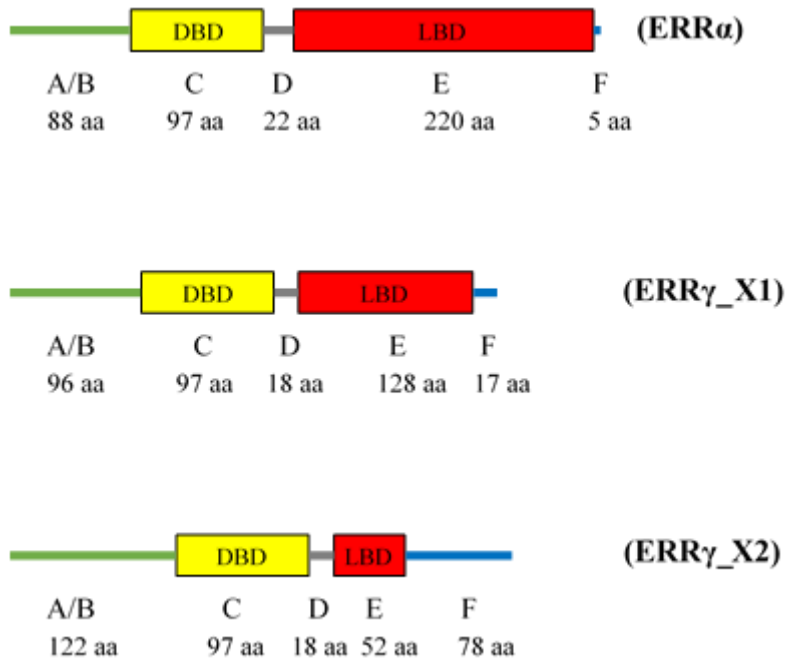


Figure 8. Secondary structure of the receptors related to *Heterotis niloticus* estrogen receptors (ERR α , ERR γ _X1 et ER γ _X2), with the five characteristic domains and the corresponding numbers of amino acids (aa).

(a)

```
HnERβ_X1 MASSPGSDLPLLFQFQEVGSSKTDGRSSPGPLSGVYAGPVPGLAMESRAVCIPSPYADSSHDYTTLGFYNPSMLGYPGSPVDPSPSVRPPPLSPA IY 94
HnERβ_X2 MASSPGSDLPLLFQFQEVGSSKTDGRSSPGPLSGVYAGPVPGLAMESRAVCIPSPYADSSHDYTTLGFYNPSMLGYPGSPVDPSPSVRPPPLSPA IY 94
HnERβ_X3 M----- 1
HnERβ_X4 MASSPGSDLPLLFQFQEVGSSKTDGRSSPGPLSGVYAGPVPGLAMESRAVCIPSPYADSSHDYTTLGFYNPSMLGYPGSPVDPSPSVRPPPLSPA IY 94
HnERβ_X1 WSPQSHPSQLPPLTLHCQQPLMYGEPTRTPWVEPKAQDHNLVESKLAGRRPLEGDEALSSSAACLTAKGDMHFC AVCHDYASGYHYGVWSCEG 188
HnERβ_X2 WSPQSHPSQLPPLTLHCQQPLMYGEPTRTPWVEPKAQDHNLVESKLAGRRPLEGDEALSSSAACLTAKGDMHFC AVCHDYASGYHYGVWSCEG 188
HnERβ_X3 -----HASLCP-----SKLAGRRPLEGDEALSSSAACLTAKGDMHFC AVCHDYASGYHYGVWSCEG 57
HnERβ_X4 WSPQSHPSQLPPLTLHCQQPLMYGEPTRTPWVEPKAQDHNLVESKLAGRRPLEGDEALSSSAACLTAKGDMHFC AVCHDYASGYHYGVWSCEG 188
HnERβ_X1 CKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACLRKCYEVGMKCGVRRERCSYRVRHRRVPOIREVMLGSGSRTQRRLESSLPPTKSF 282
HnERβ_X2 CKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACLRKCYEVGMKCGVRRERCSYRVRHRRVPOIREVMLGSGSRTQRRLESSLPPTKSF 282
HnERβ_X3 CKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACLRKCYEVGMKCGVRRERCSYRVRHRRVPOIREVMLGSGSRTQRRLESSLPPTKSF 151
HnERβ_X4 CKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACLRKCYEVGMKCGVRRERCSYRVRHRRVPOIREVMLGSGSRTQRRLESSLPPTKSF 282
HnERβ_X1 QSLALTPQLVLRIEAEPEIYLMKDMKPPFTESSMMMSLTNLADKELVLMISWAKKIPGFVELNLSQVHLLLECCWLEVLMLGLMWRSDHP 376
HnERβ_X2 QSLALTPQLVLRIEAEPEIYLMKDMKPPFTESSMMMSLTNLADKELVLMISWAKKIPGFVELNLSQVHLLLECCWLEVLMLGLMWRSDHP 376
HnERβ_X3 QSLALTPQLVLRIEAEPEIYLMKDMKPPFTESSMMMSLTNLADKELVLMISWAKKIPGFVELNLSQVHLLLECCWLEVLMLGLMWRSDHP 245
HnERβ_X4 QSLALTPQLVLRIEAEPEIYLMKDMKPPFTESSMMMSLTNLADKELVLMISWAKKIPGFVELNLSQVHLLLECCWLEVLMLGLMWRSDHP 376
HnERβ_X1 GKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATARFRELKQREYVCLKALILLNSNMCVNSPETPEELESRAKLLRLLDAVTDALVWAI SRK 470
HnERβ_X2 GKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATARFRELKQREYVCLKALILLNS-----KTPPEELESRAKLLRLLDAVTDALVWAI SRK 463
HnERβ_X3 GKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATARFRELKQREYVCLKALILLNSNMCVNSPETPEELESRAKLLRLLDAVTDALVWAI SRK 339
HnERβ_X4 GKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATARFRELKQREYVCLKALILLNSNMCVNSPETPEELESRAKLLRLLDAVTDALVWAI SRK 470
HnERβ_X1 GLTFQQQSARLAHLLMLLSHIRHLSNKGMEHLSNMKMKNVVPLYDLLLEMLDANTMHSSRVHQPPP-PGDI NPQESSSHHTAENKPLQQ 558
HnERβ_X2 GLTFQQQSARLAHLLMLLSHIRHLSNKGMEHLSNMKMKNVVPLYDLLLEMLDANTMHSSRVHQPPP-PGDI NPQESSSHHTAENKPLQQ 551
HnERβ_X3 GLTFQQQSARLAHLLMLLSHIRHLSNKGMEHLSNMKMKNVVPLYDLLLEMLDANTMHSSRVHQPPP-PGDI NPQESSSHHTAENKPLQQ 427
HnERβ_X4 GLTFQQQSARLAHLLMLLSHIRHLS-----VPV-----VWEFVFPFPG-----H 508
```

(b)

```
HnERRα TLPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGNI EYSCPATNECEITKRRRKACQACRFMKCLKVGMLKEGVRLDRVRGGRQYKRRPEVES 185
HnERRγ_X1 SLPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGNI EYSCPATNECEITKRRRKSCQACRFMKCLKVGMLKEGVRLDRVRGGRQYKRRMDAEN 193
HnERRγ_X2 SMPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGNI EYSCPATNECEITKRRRKSCQACRFMKCLKVGMLKEGVRLDRVRGGRQYKRRIDAEN 219
HnERβ_X1 G-DMHFC AVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACLRKCYEVGMKCGVRRERCS----- 187
HnERγ E-KRLC-EVCKDYASGYHYGVWSCEGCKAFFKRSIQGPTDYICPATNQCTIDKSRKSCQACLRKCYEVGMKCGVRRERCS----- 245
```

Figure 9. Alignment of the deduced protein sequences of *Heterotis niloticus* estrogen receptors (ER) beta (a), and DNA-binding domains (DBD) of ERs and ERRs (HnERβ_X1, HnERβ_X2, HnERβ_X3, HnERβ_X4, HnERγ, HnERRα, HnERRγ_X1 and HnERRγ_X2) (b). The core of the DNA-binding domain (DBD) is in boxes, and the amino acids implied in the binding with 17 β estradiol are indicated by black triangles. The little conserved domains are in gray.


```

HnERβ_X1 MASSP--GGDLP-LLOQFQEVGSSKTGDR-SSPGPLSGVYAGVPVPLAMESRAVCI P SPYADSSHDYT----TLGFYNPSMLGYPGS-VPD 81
SfERβ MAGSP--GGDLP-LLOQFQEVGSSKTGDR-CSPVSLPGVYTAGL PGLAMDSRAVCI P SPYADSSHDYA----TLSFYNPVLDYPGSSIPD 82
ObERβ MAGSP--GGDLP-LLOQFQEVGSSKTGDR-CSPVSLPGVYAGAV PGLAMDSRAVCI P SPYADSSHDYA----TLSFYNPVLDYPGSSIPD 87
OmERβ MACSPESGTDISSLQLLOQFQEVGSSKQERSSPGLLPALYS--PPSGMESRTFCIPSPYTDNNSHDYSHSHGRLAFYNPVSMGLYSRRPISD 87
OaERβ MTSPPALDADLPPLLQLEEVDSKATERVSSPGLLPAMYSS--PPVGDISHTVCI P SPYTDNHDYSHSHGRLAFYNPVSMGLYSRRPITD 87
CcERβ M--SVYAEGLDSP-LLOQFQEVGSSRVGGHVLSPA----FNSSSSLPVESHPICI P SPYTDLGHDFI----TLPFYSPALLGYGAAPLSD 78

HnERβ_X1 SPS-VRPPLSPAIIYWSPHGHPGQLPPLTLHCOOPLVYGEPTRTPWVEPKAODHNLVSS-----KLAGRRRLEGE-ALSSSAACLT 162
SfERβ SPS-VRPPLSPAIIYWSPHGHPGQLPPLTLHCOOPLVYGEPTRTPWVEPKAODHALVSS-----KLTGRRLLEGE-ALSSSAACLTG 163
ObERβ SPS-GRPPLSPPIIYWSAHSHPGQLPPLTLHCOOPLTYGEPTRTPWVEPKAODHALVSS-----KLTGRRLLEGE-ALSSSAACLTG 163
OmERβ SPS-LCPPLSPSLFWPNHGQON-MPSLTLHCQPLVYSEHNTHTPWVEPKP--HGLSPSSPLLHPTKLLGKRLLEDGEE-VNSSSASCVVV 172
OaERβ SPSLLOPPLSPSAFWPSSHHS-VPSLTLHCQPLVYNEPSPHAPWLPKW--HSTSPGSSIIITCNKLLGKKSSEDGAEVKSSSCSALG 174
CcERβ CPS-VRQSLSPSLFWPPHSH--VSSVALHCOQTRLQONHPTGGSWAELTPHDHSEEENC-----KPLAKRVADAE-----TSASLRG 153

HnERβ_X1 KQDMHFCVACHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCVEVGMKCGVRRRERCSY--RGVR 250
SfERβ KSDMHFCVACHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCVEVGMKCGVRRRERCSY--RGVR 251
ObERβ KSDMHFCVACHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCVEVGMKCGVRRRERFNY--RGVR 251
OmERβ KADMHFCVACHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCVEVGMKCGVRRRERCSY--RGAR 260
OaERβ KADMHFCVACHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCVEVGMKCGVRRRERCGF--RGAR 262
CcERβ KADMHFCVACHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCVEVGMKCGVRRRERDQY--RGAR 263

HnERβ_X1 HRRV-POIREVM---L---GSGSRTORRLE----SSLPP---TKSFQSLALTPEQLVLRITAEAPPEIYLMKDMKKPFTESSMMMSLTN 325
SfERβ HRRV-POIRDTA---M---GPCPRAORRLE----SSLSP---MKNFQPLALTPEQLVSRIMDAEPPEIYLMKDTKKPFTESSMMMSLTN 326
ObERβ HRRV-POIRDTA---M---GPSRPTORRLE----GSLPP---MKNFQPLALTPEQLVSRIMDAEPPEIYLMKDTKKPFTESSMMMSLTN 326
OmERβ HRRV-POGRGVSGGLV---GVGTRACMRLE----GSSHP---OLEVHHSLLTPEQLISGIMEAPPEIYLMEDLKKPFTESSMMMSLTN 338
OaERβ HRRGGPOPRDITGOSLVRVGLGSRGQRHLHLGLTSLTLPVPHNTHHSSIMSPKEFISRIMEAPPEIYLMEDLKKPFTESSMMMSLTN 352
CcERβ DKRLA-----RFSGRMRTG--PRSQE-----IKSVPRPLGGNVVSIALSPEELIARIMDAEPPEIYLMEDVKKPFTESSMMMSLTN 319

HnERβ_X1 LADKELVLMISWAKKIPGFVELSLSDQVHLLCCWLEVLMLGLMWRSDVHPGKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATRFREL 415
SfERβ LADKELVLMISWAKKIPGFVELSLSDQVHLLCCWLEVLMLGLMWRSDVHPGKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATRFREL 416
ObERβ LADKELVLMISWAKKIPGFVELSLSDQVHLLCCWLEVLMLGLMWRSDVHPGKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATRFREL 416
OmERβ LADKELVLMISWAKKIPGFVELSLSDQVHLLCCWLEVLMLGLMWRSDVHPGKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATRFREL 428
OaERβ LADKELVLMISWAKKIPGFVELSLSDQVHLLCCWLEVLMLGLMWRSDVHPGKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATRFREL 442
CcERβ LADKELVLMISWAKKIPGFVELSLSDQVHLLCCWLEVLMLGLMWRSDVHPGKLI FSPDLKLNREDEGSCVEGIMEIFDMLLAATRFREL 409

HnERβ_X1 KLQREEYVCLKALILLNSNMCVSNPETPEELESRAKLLRLLDAVTDALVWASRKGLTFQOQSARLAHLMLLSHIRHLSNKGMEHLSTM 505
SfERβ KLQREEYVCLKALILLNSNMCVSNPETPEELESRAKLLRLLDAVTDALVWASRKGLTFQOQSARLAHLMLLSHIRHLSNKGMEHLSTM 506
ObERβ KLQREEYVCLKALILLNSNMCVSNPETPEELESRAKLLRLLDAVTDALVWASRKGLTFQOQSARLAHLMLLSHIRHLSNKGMEHLSTM 506
OmERβ KLQREEYVCLKAMILLNSNMCSNPERAEDLESRGKLLRLLSDVTDALVWASRKGLSFQOQSARLAHLMLLSHIRHLSNKGMEHLSTM 518
OaERβ KLQREEYVCLKAMILLNSNMCSNPERAEDLESRGKLLRLLSDVTDALVWASRKGLSFQOQSARLAHLMLLSHIRHLSNKGMEHLSTM 532
CcERβ KLQREEYVCLKAMILLNSNMCLSSSEGGEELDSRSKLLDLDSDVTDALVWASKTLGSLFQOQSARLAHLMLLSHIRHLSNKGMEHLSTM 499

HnERβ_X1 MKMNVVPLVDLLLEMLDANTMHSSRVHQRPPGDINPOES-----SSHTAENK 554
SfERβ MKMNVVPLVDLLLEMLDANTMHSSRVLHNRPOGEASPOE-----SPSSAENK 555
ObERβ MKMNVVPLVDLLLEMLDANTMHSSRVHNRPOGEASPOE-----SPSSMAEKK 555
OmERβ MKMNVVLDLLEMLDANTMHSSRMSATHDPSNNDPTEPPAAPA-----AVDTQFLLTFQNPPEESQTLLESSTSSQGAGC 595
OaERβ MKMNVVLDLLEMLDANTMHSSRMSATHDPSNNDPTEPPAAPA-----AVDTQFLLTFQNPPEESQTLLESSTSSQGAGC 622
CcERβ MKMNVVPLVDLLLEMLDANTMHSSRVLHNRPOGEASPOE-----SPSSAENK 555

HnERβ_X1 -----P-----LQ 558
SfERβ -----P-----LK 559
ObERβ -----P-----LQ 559
OmERβ -----PREGRCPVQ 604
OaERβ AGSHLDSNDYISVEHWSLNNGDAGPGPSAEPPTYVIPDRVVTETA 667
CcERβ -----GGTLGT 544

```

Figure 9c. Alignment of the deduced protein sequences of *Heterotis niloticus* estrogen receptors (ER) beta (HnERβ_X1) with ERβ of *Scleropages formosus* (SfERβ), *Osteoglossum bicirrhosum* (ObERβ), *Oncorhynchus mykiss* (OmERβ and OmERγ), *Oreochromis aureus* (OaERβ and OaERγ) and *Cyprinus carpio* (CcERβ and CcERγ). The core of the DNA-binding domain (DBD) is in boxes, and the amino acids implied in the binding with 17 β estradiol are indicated by black triangles. The little conserved domains are in gray.


```

HnERγ M-----L LSLVYSS-----ATPNIIPSP LMDNCOQDYS-----27
OaERγ M-----MAAASSPEK-----LLOLQEVDSRRAG--SRI LSPILGSSSPGLSHETSQPICIRSPYTDLGHDFTTI-----62
OmERγ MSQYRRLPGLPSEL PQSPMAASPLPERDSATLLK LQEVDRSRVGRGGRILSPIFSAPSPALPME-AHPICIPSPYTDIGHDFNPL-----84
CcERγ MSS-----SPGPAPASASPAL--ESGKANRGDPSNTRSRLYTSP--LGM D-NQTVCI PSPYVEACQDYSPLHGGEISH-----68
HnERγ ----VQOYAC PGLSDSSF--LDP PFDWKPDSNFLSVKKITPPSSPLAEPEP-----QSLFSTDKQQQRKRKTNGM-----93
OaERγ ---PFYSP---TIFSYGGPSISECSSVHQSLASLFWPSSHGRVGTPI TLHC-PQGRSQGQSAOTPW---DSVITTSKSVRR-----RSQ137
OmERγ ---SFYSP---TLLSYAGPALSDCPSTHQSLSPSLFWPPQAHMGFPLSLHHRPQSRPQOGQPTRVSW--AEPHALSESSKPLRK-----RSQ163
CcERγ GALTLIYSPMSSTVLGYHPVPSSESLV--PLSPTIFWPPYITHTALS LHCPPMAYSETHAHTAWEDAKHTHLNQSSVLT HAKLLGQOL-----155
HnERγ CRVKSLOPFPG--DETEKRLCFVCKDYASGYHYSVWSCEGCKAFFKRSIQGPTD YICPANNQCTIDKSRRKSCQACRLRKC YEVGMVKNVGR-----183
OaERγ ESEESMVSSGG---KADLHYCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKC YEVGMTKCGTR-----226
OmERγ EGEEETVLSLEGG---KAE LHFCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKC YEVGMTKCGMR-----252
CcERγ DGGDDLNRSPGFVYKGDTHFCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNNYICPATNQCTIDKSRRKSCQACRLRKC YEVGMVKGCVR-----247
HnERγ RRRRPSQEVQGI RMCQHL SVMGEEAS---L L SQEKSHLSASLEET-----QQLLFS PKQLISRLEVEPPDI YLWMDLKKPTEV-----260
OaERγ KERGNVYRNSQARRLTR-LSSQGKTAEPKGI TGP AEGSLNKPEKP-----AL TPEQLIERILEAEPEPI YLMKDEKRPLTEA-----301
OmERγ RDRSSYRGHKPRRVGR-FFTRGTASGPKRVL--AEGS--EPIKELC-----PTVLTPEQLIGRI MAEPEPIELQKDMRRPLTEA-----327
CcERγ RERQSYRGARHRNPOI--RDS SDGALGVRRRSQHHLEFPLNRTHHFFPSGGDAEGRGLGLSPEQLVNCI LEAEPPQI YLREPIK KPYTEA-----336
HnERγ SMMTSLTLADKELF YMI SWAKKIPGFTELGLSDQVHLLCCWLEVLMLGLIWRSDVHPGKLI FSPDLILSRDEGSCIEGILEIFDMLLAAT-----352
OaERγ SVMMLLTNLADKELVHMI SWAKKIPGFVELSLVDQVHLLCCWLEVLMLGLMWRSVDHPGKLI FCPDLSLREEGSCVQGFVEIFDMLLAAT-----393
OmERγ NVMMSLTNLADKELVHMI SWAKKIPGFVDLQLFDQVHLLCCWLEVLMLGLMWRSVDHPGRLI FSPDLSLREEGSCVQGFVIFDMLLAAT-----419
CcERγ SMMMSLTNLADKELV LMI SWAKKIPGFVELT LSDQVHLLCCWLDLMLGLMWRSVDHPGKLI FSPDLKLNREEGNCVEGIMEIFDMLLAAT-----428
HnERγ SRFRELRLOKEEYVCLKALVLLNSSIYLYTPNNGGYEHQSRRAKLDQLLN AVTDALTWVIAKLG LSFQQQSARLAHLLMLLSHIRHVS NKGM D-----444
OaERγ TRVRELKLOREEYVCLKAMILLNSNMCLSSSDGSEDLQSRSKLLHLLDAMTDALVLAIGKTGLTFRQYTRLAHLLMLLSHIRHVS NKGM D-----485
OmERγ SRFREKLQREEYVCLKAMILLNSNMCLSSSEGSEELQSRSKLLRLLDAVTDALVWAIAKTGLSFQQQSARLAHLLMLLSHIRHVS NKGM D-----511
CcERγ SRFRELKLQSEYVCLKAMILLNSNCSL PQTPEDVESRQKVLRLLD SVTDALVWTISRTGLSSQEQSIRLAHLLMLLSHIRHVS NKGM D-----520
HnERγ LHCMMKMKVVPFYDLLLEMLDAHVMHNSRI LSPSEPEP-----RAGGESQ-----557
OaERγ LHCMMKMNIVPLYDLLLEMLDAHIMHSCL--PHOPPPQD SKDQSEAPALHSSAGGPSNTWTIPSSA-----RAGGESQ-----557
OmERγ LHCMMKMNIVPLYDLLLEMLDAHIMHSRL--PHO-----ANSAGPCPEVSP-----QPT-----560
CcERγ LSTMKRKNVVL YDLLLEMLDANTSQSSRMLAAHTEASLOS D TQOTTEILHTSRQSNQESRHSPCAEETVDKTLHC SLHRADMDTD-----606

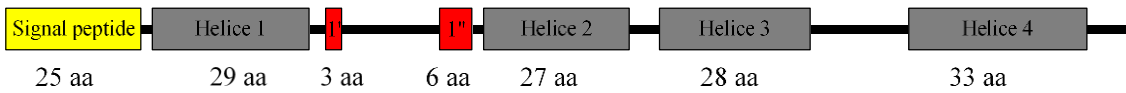
```

Figure 9d. Alignment of the deduced protein sequences of *Heterotis niloticus* estrogen receptors (ER) gamma (HnER γ) with ER γ of *Oncorhynchus mykiss* (OmER γ), *Oreochromis aureus* (OaER γ) et *Cyprinus carpio* (CcER γ). The core of the DNA-binding domain (DBD) is in boxes, and the amino acids implied in the binding with 17 β estradiol are indicated by black triangles. The little conserved domains are in gray.

(a)

```
1 M M R Q S Q G S R F L W S A V L L S L C A Q C H C V N F S D L M D R A
1 ATGATCGCTCAATCCCAAGGTAGCAGATTCCTGTGGTCAGCCGCTGCTGTCCCTCTGTGCCAGTGTCACTGTGTCAACTTCAGCGACCTGATGGACAGGGCA
36 V Q L S D K L H F L S N S L K I D L D S H F P A V G K R I P R L S L
106 GTTCAGCTCTCAGACAACTGCACCTCCCTCAGCAACTCCCTCAAATCGACCTGGATTCTCATTTCAGCTGTGGGAAAAGGATCCCTCGCCTGTCTGTGC
71 H T S S L R T P M D K S Q V L S L P E S E L L S L I Q S L L L S W K N
211 CACACTTCTCCCTCAGAACCCATGGACAAGAGCCAGGTGCTCAGCTGCCGGAGTCGGAGCTGCTGTCCCTCATTTCAGTCCCTCCTCCTGTCTGGGAAGAT
106 A L P V L S S E A S S L P H A W Q N G I S T K A K Q L Q E H W H S L H
316 GCACTGCCGGTCTCTCTCCGAGGCCTCTCTCTCCCCACGCTTGGCAGAAGCCATCAGCACCAAGCCAAACAGCTACAGGAGCACTGGCACAGCCTCCAT
141 T G L S T L A H K M G S S S D S A A V P Y D E S S L G D D N S R L A G
421 ACTGGACTCAGCACACTGGCACACAAGATGGGCTCTCTCCGACTCTGCTGCCGTCCGTACGACGAGAGCAGCCTTGGCGACGACAATTCTCGACTCGCCGGC
176 F H F L L S C F R R D S H K I D N F L K I L R C R V H K Q K L D V *
526 TTCCACTTCTGCTGTCTCTCAGGAGGACTCGCACAAAATCGACAATTCTCTCAAATTTCTCCGCTGCAGGGTCCACAAAACAAAACACTGGATGTGTGTGA
```

(b)



(d)

```
HnPRL M - MRQSQGS RFLWSA - VLLSLCACCHCVNFS DLMDRAVQLSDKLHFLNSL KIDLDSHFPAV GKRI - PRLSLCHTSSLRT 77
SfPRL M - LRRRAQGS KLLWSA - VLLSLFVHCHCVDFGDLMDRAVQLSDKLVHFLSTLSL KSDLDYHLPVAVGKKI - PHLSLCHTSP LKT 77
CcPRL M - - - TQGSRL YFAV T VLMCAFVS INGVGLNDL LERASQLSDKLHSLSTSL TNDLDSHFPPVGRVMMPRPSMCHTSSLQI 76
OmPRL MA - RRSQGT KHLAV - - - LCLVVSCHAI GLSDLMERASQSDKLHSLSTSL TKDLDSHFPPMGRVMMPRPSMCHTSSLQT 76
OnPRL MRQRR TSGTNLFMTV - - - LCVVAMCRAVP I N E L F E R A S O H S D K L H S L S T L Q E L D S H F P P I G R V I M P R P A M C H T S S L Q T 77

HnPRL PMDKSQVLSLPESELLSL IQSLLL SWKNAL P V L S S E A S S L P H A W O N G I S T K A K Q L Q E H W H S L H T G L S T L A H K M G S S S D S - 156
SfPRL PMDKNQVLSLSESELLSL VQSLLL SWTDAL L L L S S E A S S L P H P S T S T I Q V K A K Q L R E N S D T L G T G L S R L A H K M G S S S - S - 155
CcPRL PNDKQAL K I P E D E L L S L A R S L L L A W S D P L A L L S S E A S S L A H P E R N T I D S K T K E L O D N I N S L G A G L E H V F Q K M G S S S D N L 156
OmPRL PKDKEQALKVSENEL I S L A R S L L L A W N D P L L L L S S E A P T L P H P S N G D I S S K I R E L O D Y S K S L G D G L D I M V N K M G F S S Q Y I 156
OnPRL P I D K D Q A L Q V S E S D L M S L A R S L L Q A W S D P L V V L S S S A S T L P H P A Q S T I F N K I Q E M Q Y S K S L K D G L D V L S S K M G S P A Q A I 157

HnPRL A A V P - - Y D E S S L G - D D N S R L A G F H F L L S C F R R D S H K I D N F L K I L R C R V H K O K L D V C 209
SfPRL S Y I P L R Y T E N S L G G D D S R L V N F H F L L S C F R R D S H K I D N F L K I L R C R A N K O K P D M C 211
CcPRL S S L P F - Y T S - S L G D K T S R L V N F H F L L S C F R R D S H K I D S F L K V L R C R A A K R P E M C 210
OmPRL S S I P F - K G G - D L G N D K T S R L I N F H F L M S C F R R D S H K I D S F L K V L R C R A T K M R P E A C 210
OnPRL T S L P Y - R G G T N L G H D K I T K L I N F N F L L S C L R R D S H K I D S F L K V L R C R A A K M O P E M C 212
```

(c)

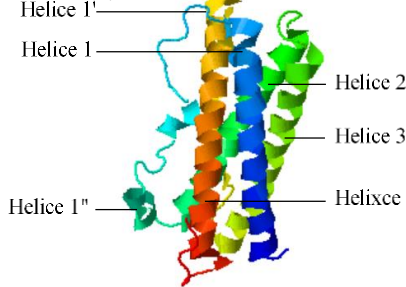


Figure 10. Nucleotide sequences and deduced protein sequences (a) of *Heterotis niloticus* prolactin (HnPRL), secondary structure of HnPRL with the different domains (b), tertiary structure obtained with I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) showing the α -helices of HnPRL (c), and alignment (d) with the PRLs of *Scleropages formosus* (SfPRL), *Oncorhynchus mykiss* (OmPRL), *Oreochromis niloticus* (OnPRL) and *Cyprinus carpio* (CcPRL). In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, the four major helices (1, 2, 3 and 4) are laid out successively in boxes, and the short helices 1' and 1'' are double-underlined. In figure (d), the four conserved cysteines are in boxes, the little conserved domains are in gray.

(a)

```
1 M M W K D S G A P L I L S L L V S V V L D S A C L S P P G K P R L T S
1 ATGATGTGGAAAGACTCTGGAGCCCCTGATTCCTGCTACTGCTGGTCTCTGTGGTCTGGACAGTGCCTGTCTGTCTCCCCAGGAAAGCCCTCGGCTCACCAGC
36 C R S P E K E T F T C W W E P G L D G G L P T N Y S L F Y R K E N S D
106 TGCCGATCTCCTGAGAAAGAAACCTTACCTGTGGTGGGAGCCAGGCTTGGATGGTGGCTGCCACCAACTACTCCCTCTTCTACCAGCAAGAAATCTGAT
71 S V Y E C P D Y H T A G K N S C F F S K N D T S I W V N Y N I T V V A
211 TCTGTGTATGAGTGCCAGACTATCACACTGCAGGAAAGAACTCCTGCTTTTCAGCAAGAATGACACTTCAATCTGGGTGAATTACAACATCACTGTGGTGGCA
106 T N A L G S N F S D P V D V D V V Y I V Q P H T P E N L T V V L E G E
316 ACCAATGCCCTAGGAGCAACTTCTCAGACCCCTGGATGTGGATGTTGTACATCGTCCAGCCCCACACCAGAGAACTTAACCGTTGTTTTGGAAGGTGAA
141 E N P F L V V R W E P P R M A D T R S G W I T L I Y Q L R V K I E K E
421 GAAAACCCATTCTGGTGGTGGGAGCAACCCCTCGCATGGCGGACACAGCCTCTGGCTGGATTACTCTCATTACCAGCTCCGTGTAAGCTGGAAAAAGAA
176 E E W E E E F A G Q Q K Q F N I F S L E S G E V Y M V Q V R C K P D H
526 GAGGAGTGGGAGGAGCACTTTGCTGGACAGCAGAAGCAGTTAAACATCTTCAGCCTGCATCGGGGGAGGTGTACATGGTACAGGTCGGCTGTAAAGCTGACCAT
211 G F W S E W S T T A Y V K I P D Y I S K D R S V W I L I T I F S A F L
631 GGCTTCGGAGTGAATGGAGCACCCCGCTTATGTGAAAATCCCAGCTACATTTCTAAAGACAGTCAAGTTGGATCCCTGATACTATTTCTCAGCATTCCCTT
246 F L V F I W T M H I N R N S V K E F L L P P V P G P K I K G F D T Q L
736 TTCTTGTCTTCATATGACTATGCACATTAACAGAAACAGTGTGAAACATTTTCTCCTCCCCCTGTGCTGGACAAAGATCAAAGGATTGATACTCAGCTG
281 L K S G K S E E V F N A L V I Q G F T P P S Y Y D D L L V E Y L E V Y
841 CTCGAAGTGGCAATCAGAGGAGGCTTCAATGCCCTGGTTATCCAGGGCTTCACACCACCGTCATACTATGATGATCTGCTGGTGGAGTACCTGGAAGTGTAC
316 D N K E Q V L K L D R K D V Q E S R F K S K S S S D S D S G R G S C D
946 GATAACAAGGAGCAAGTGTAAAGCTTGTAGAAAGGATGTTCAAGAGAGTCGGTTTAAAGTCCAAGAGCTCTTCCAGACAGTACTCTGGGAGAGGCAGTTGTGAC
351 S R T L L M E K C G E P K E D Q T M I S Q V K A G E A S R E I L M K N
1051 AGCCGCACATTATTAATGGAGAAATGTGGTGAACCCAAAGAGGACAGCACCATGATATCACAAGTAAAAGCTGGTGAAGCTAGCAGAGAAATATTAATGAAAAAT
386 D S S I P E S A N V K T K T W P T I C S S Y H P E W N L C Y E S I P E
1156 GACAGCAGTATCCCAGAATCTGCTAACGTGAAAACCAAAACATGGCCCACTATCTGCTCATCTATCACCCCGAATGGAATCTATGCTACCACAGTATTCCAGAG
421 M P V Q C Q V F S S H F S S T H Q Q D Y W E H P A E I Y H K K P P S G
1261 ATGCCAGTCCAGTGCCAGGTACCTAGCAGCCACTTTTCTCCACACACAGCAGACTATGGGAACACCCCTGCAGAGATCTACCACAGAGGCTCCGTCTGGC
456 Y R Y T Q A Y S E F N I K G I G M K A S A F P P S K S M E Y V E V Q E
1366 TACAGATATACCAAGCTTACAGTGAAGTTAACATTAAGGAATGGCATGAAAGCATCCGCTTTTCTCCGTCTAAGTCCATGGAGTATGTGGAGGTTTCAGCAT
491 V N Q E N T L I L K P I S Q E G P D H M E C G G H D Y S K V N R V V S
1471 GTCACCAAGAGAACACATTGATATTAACCCATCAGCCAAGAGGGTCCAGACCACATGGAGTGTGGAGGACATGATTACAGCAAGTCAACCCGGTGGTCCAGT
526 D S V L L L Q R D M S G Q C L S D C Q D R E T P G Q C T Q Q S G K
1576 GACAGTGTCTCTGCTTCAAAGGGACATGTCTGCCAGTGCCTCAGTACTGCCAGGATAGGGAGACACCAGGGCAGACTGCACACAGCAACAGTCTGGAAAAG
581 P A E H L Q V P V A P E T R L T L N G Y V D T A T M L S Y *
1681 CCAGCAGACACCTGCAAGTCCCTGTGGCTCCAGAAACAGCCTGCACACTGAATGGCTATGTGGACACAGCCACCATTGTATCATACTAA
```

(b)

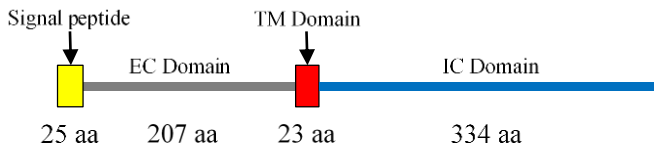


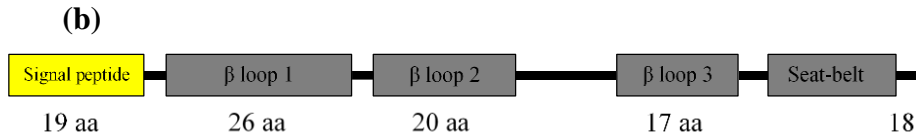
Figure 11. Nucleotide sequences and deduced protein sequences of (a) *Heterotis niloticus* prolactin receptor (HnPRLR), and box modeling with the different domains (b). In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, the transmembrane domain (TM) in a box located between the extracellular domain (EC) in the N-terminal region and the intracellular domain (IC) in the C-terminal region.

(a)

```

1 M M A C V L L L V V T M V W V L V G A G P Y CDLTLNVSISLENE
1 ATGATGGCCTGCGTGTCTGCTCGTGTGACAATGGTCTGGTGGTGGGAGCCGGCCCTACTGCGACCTAACCAACGTTCCATCTCCCTGGAGACGAG
36 E C G K C I T I T T T S C T G W C F T Q D P V Y K S S L A P Y I Q H T
106 GAGTGTGAAAATGCATCACCATCACCACGACCTCGTGCACCGGCTGGTGTTCACCCAGGATCCCGTATACAAGAGCTCCCTGGCCCCGTATATTCAGCACAGC
71 C N F K E V T H E S A V L P G C P E G V D P H F S Y P V A L S C E C S
211 TGCAACTTCAAGGAGTGACGCATGAATCAGCTGTCTGCCAGGCTGCCCGAAGGAGTGGACCCCATTTTCAGTACCCGGTGGCCCTGAGTTCGCAGTGCAG
106 R C N T D T T D C G A L N E D V S G C L G L *
316 CGCTGCAATACGGACACACAGGACTGCGGTGCCTCAACGAAGACGTTTCGGGCTGTCTGGGGCTCTAA

```



(c)

```

HnFSHβ MMACV L - - - L LVV T M V W L - V G A G P Y C D L T L N V S I S L E N E C H C I T I T T T S C T G W C F T Q D P V Y K S S L A P Y I Q H T C N F K E V T H E S A V L P G C P E G V D P H F S Y P V A L S C E C S 81
AgFSHβ M - T C M L - - - L LVV T M A W P L - V G A G P H C D - - - L N N I T I P L E N E E C E H C I T I T T T S C T G W C F T Q D P V Y K S S L A P Y I Q H T C N F K E V A Y E S A 80
CcFSHβ M R M H F V - - - V M V M L L P - A L - M M A G S E C R S S C R L T N I S I T V E S E E C G S C I T I T T T A C A G L C K T Q E S V Y R S P L M L S Y O N T C N F E W T Y E T Y 84
OmFSHβ M Y C T H L K M L Q L V V M A T L W V T P V R A G T D C R Y G C R L N N M T I T V E R E D C H G S I T V T T - C A G L C E T T D L N Y Q S T W L P R S Q G V C N F K E W S Y E K V 88
OnFSHβ M - - - - - Q L V V M A A V L A L - A G A E Q D C S S G C R P K N I S L P V D T - - - - - Q G F V D T T I C E G C C F Q K D P N F I H T D D W P K O K T C N - G E W S Y E V K 75

HnFSHβ V L P C P E G V D P H F S Y P V A L S C E C S R C N T D T T D C G A L N E D V S G C - - - L G L 127
AgFSHβ F L P C P E G V D P H F T Y P V A L S C E C S R C N T D T T D C G A L S E E V S G C - - - Q G H 126
CcFSHβ E F K G C P A R A D S V F T Y P V A L S C E C S K C N S D I T D C G A L S Q O T L S C - - - N A H 130
OmFSHβ Y L E G C P S G V N P L F - I P V A K S C D C I K K T D N T D C D R I S M A T P S C I V N P L E M 137
OnFSHβ Y T E C P R G - - - F I Y P V A R K C E C T A C N A N - T D C G T L S G Y I P S C - - - - - 113

```

Figure 13. Nucleotide sequences and deduced protein sequences of (a) the beta subunit of *Heterotis niloticus* follicle-stimulating hormone (HnFSH β), box modeling of HnFSH β with the different domains (b), and alignment (c) with *Arapaima gigas* FSH β (Ag HnFSH β), *Oncorhynchus mykiss* FSH β (Om HnFSH β), *Oreochromis niloticus* FSH β (On HnFSH β) and *Cyprinus carpio* FSH β (Cc HnFSH β). In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, β -loops 1, 2 and 3 and the “seat-belt” are laid out successively in boxes. The predicted N-glycosylation site is double-underlined. In figure (c), conserved cysteines across species are in boxes, the little conserved domains are in gray.

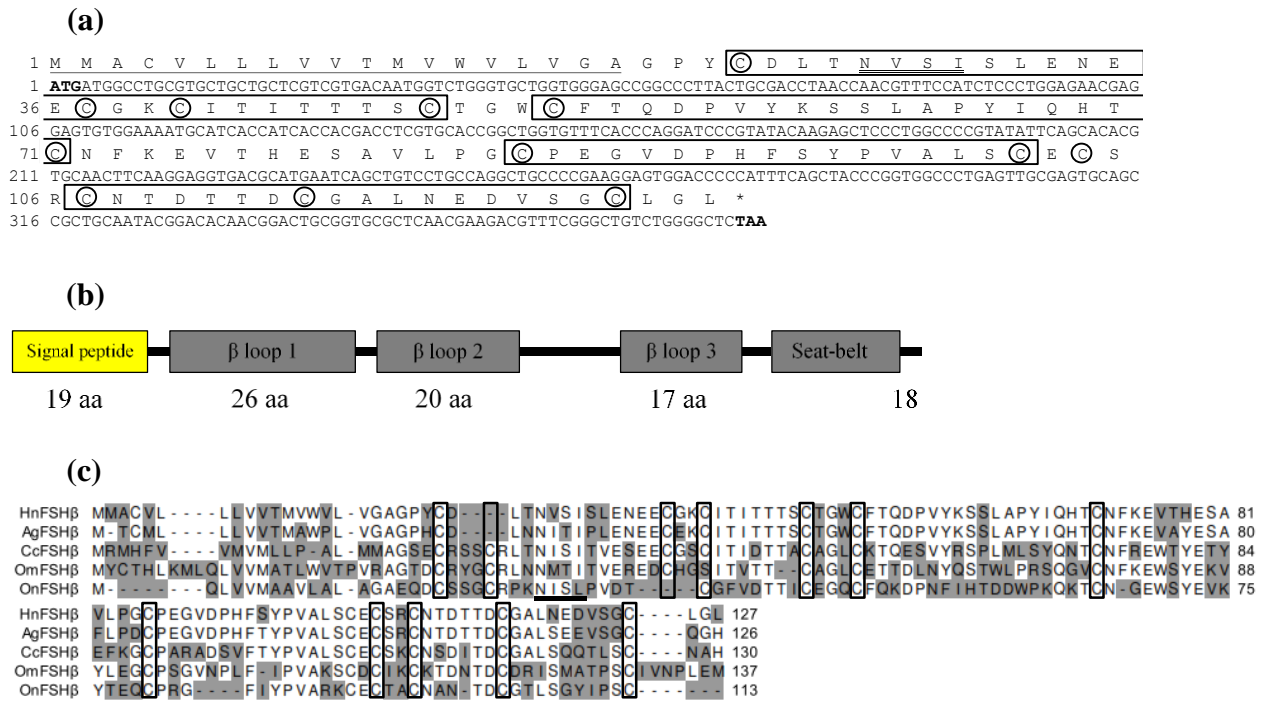


Figure 13. Nucleotide sequences and deduced protein sequences of (a) the beta subunit of *Heterotis niloticus* follicle-stimulating hormone (HnFSH β), box modeling of HnFSH β with the different domains (b), and alignment (c) with *Arapaima gigas* FSH β (Ag HnFSH β), *Oncorhynchus mykiss* FSH β (Om HnFSH β), *Oreochromis niloticus* FSH β (On HnFSH β) and *Cyprinus carpio* FSH β (Cc HnFSH β). In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, β -loops 1, 2 and 3 and the “seat-belt” are laid out successively in boxes. The predicted N-glycosylation site is double-underlined. In figure (c), conserved cysteines across species are in boxes, the little conserved domains are in gray.

(a)

```

1 M L L W V Y H A R Q R T P T M H Q V W M L L L G L L F V L L D Q K L V
1 ATGTTACTCTGGGTTTACCACGCCAGACAGAGAACCCCAATGCACCAAGTGTGGATGTTGCTCTTGGGGCTTCTCCAGTCTTTGGACCAGAAGTTGGTG
36 E A ▼H G C F S N R T S C S F F C L G A K I E Q M P S T I P W N T S S I
106 GAGCGCATGGCTGCTTTAGTAACAGGACCAAGTGTCTCTTCTTGCTCTGGGGGCCAAAATCCATCAGATGCCAGACCATTCCCTGGAAACACCAGCAGCATT
71 E F R L T Q L R T F P P K A F W G L T E V S R I M L S E N G L L E E I
211 GAGTTTAGGCTGACGCAGCTGAGGACCTTCCCACCAAGGCCTTCTGGGACTGACAGAGGTGCTCGAATCATGCTGTCAGAGAACGGACTACTGGAAGAAATT
106 G P Q A F Y N L S K L V E I T I T K S K H L V V I H K D A F L D L P K
316 GGACCTCAAGCCTTTTACAACCTCTCCAACCTGGTAGAATAACAATCACCAAGTCAAAACATCTGGTTGTATTATAAGGATGCGTTCTTAGATCTGCCAAAA
141 L E Y L T I M N T G L K L L P D F S K I N S A A H G F L L D L Q D N M
421 CTGGAATACCTAACCATATGAACACTGGTCTCAAGTTGCTGCCTGATTCTCCAAAATCAACTCAGCAGCAGTGGCTTCTCTGCTGGACCTCAAGACAACATG
176 H I D V I P F N A F L G L S A D T I K E L R L T K N G I T E V R S H A
526 CACATAGATGTGATCCACCCAATGCCTTTTGGGACTGAGCGTGTACTATAAAAGAGCTGAGACTCACAAAATGGGATCACTGAGGTCGGGAGTCATGCC
211 F N G T K M Q K L S L M G N Q L L R Q I H R H A F M G A E G P I V L D
631 TTTAATGGCACAAAATGCGAAGCTGTCTCTATGGGAACCCAGTGTGTTAAGGCAGATCCACAGGCATGCCTTTATGGGAGCGGAGGCCCCATTGTACTGGAC
246 I S R T A V S T L P E N M L R G I K L L T A V G V E N L K K L P R I E
736 ATCTCAGCGGCTGTACAGCACTTGCCCGAGAACATGTACGTGGCATAAAGCTGTTGACGGCTGTGGGAGTGATAACCTGAAGAAGCTGCCAGSTGGAG
281 I M G F T Y L R V L I W V I S V L A I A G N L L V L A V L L S S H Y K
841 TTATTTACCAACTCATCGAGGCCAACCTGACATACCAAGCCACTGCTGTGCCTTCGCCAACCTCAGGAAGAACATGTCCTGGAAAACCTCCATGTGCAGTTTA
316 P H I Q E E D E P E F F L D H C W D V S K V S C F P K P D A F N P C E D
946 CCCCACATCCAGGAGGACGACCACCTTTCTTCTGGACACTGCTGGGATGTCAGCAAGGTGTCTTGTCTTTCCCAAGCCAGATGCCTTCAACCCATGTGAGGAT
351 I M G F T Y L R V L I W V I S V L A I A G N L L V L A V L L S S H Y K
1051 AITACGGTTCACCTACCTTCTGGTGTCTCATCTGGGTTATTAGTGTGTGGCAATTCGGGGCAACCTGTGTCTGCTGGCTGTCTGTGAGCAGCCACTACAAG
386 L T V P R F L M C H L A F A D L C M G I Y L L L I A I V ▼ D T R T R S H
1156 CTCACCGTGGCCCGTCTCTCATGTGCCACTTAGCCTTTTGTGACCTGTGCATGGGTATCTACCTGTACTCATTTGCCATTTGGGACACCCGCACCTGCGCAC
421 Y Y N E G I E W Q T G P G C A T A G F F T V F A S E L S V Y I L T A V
1261 TACTACAATCATGGAATGGAATGGCAAACAGGGCCCTGGCTGTGCAACTGTGGCTTCTTACCCTCTTCGCCAGGAGTATACAGTCTACAGCTTACTGCAGTC
456 T L E R W H T I T F A I R V D R R L R L R H A C A I M A V G W L F A S
1366 ACTCTTGAGCGCTGGCACACCATACCTTTGGCTACGGTTCACCGACCCCTCCGACTGAGGCACGCTGGCCAAATCATGGCCGTTGGCTGGCTCTTTGCCTGG
491 L A A V L P V V G V S S Y S K V S I C L P M D V E T T G A Q V Y V V V
1471 CTAGCTGCCGTGCTGCCTGGTGGGTTGAGCAGCTATTCAAAGTTCAGCATATGCTGCCAATGGATGTGGAGACAACAGGTGCCAGGTCATGTGGTGGTG
526 L L L N V L A F L A V F V C Y L R I Y I L T V R N P M N F V P A S A D M
1576 CTGTGTGCTCAATGTCTGGCCCTCTCTGGCAGTATTGCTGTCTACCTGCCATCTACCTGACAGTGCCCAATCCCAACTTTGTGCCAGCCAGCCGACATG
561 R V A K R M A V L I F T D F L C M A P I S F F A I S A A L R L P L I T
1681 CGCGTTGCCAAACGGATGGCTGTGCTACTCTTCCAGACTTCTGCAGGCCCCCATTTTCATTCTTGCATCTCTGCAGACTCAAAGTCCATCACTCACT
596 V S H A K V L L V L F Y P I N S C S N P F L Y A F F T K T F K R D F E
1786 GTGTCACATGCCAAGTACTGCTAGTGTCTTCTACCCATAAACTCATGCTCCAACCCATTCTCTACGCCTTCTACCAAGACTTTCAAGAGGACTTCTTCT
631 I L A S R F G C F K A R A Q M Y R T E T S S V Q N G G W G P T H K T S
1891 ATCTTGGCCAGCGCTTTGGTTGCTTCAAAGCCAGGCACAGATGTACCCGACTGAGACGCTCCTCAGTCCAGAATGGGGGTGGGGGCCACCCCAAGACCAGT
666 D A T L Y S L V P V N H V H *
1996 GATGCTACTCTACTCCCTGGTGGCAGTCAACATGTGCACTGA

```

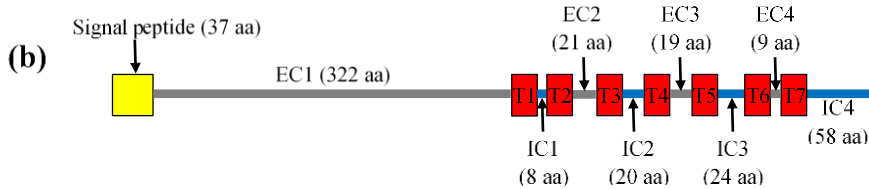


Figure 14. Nucleotide sequences and deduced protein sequences (a) of *Heterotis niloticus* follicle-stimulating hormone receptor (HnFSHR), and box modeling (b) with the different extracellular domains (EC1 to EC4), transmembrane domains (T1 to T7) and intracellular domains (IC1 to IC4) of HnFSHR. In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, T domains are in boxes, IC domains are double-underlined, and EC domains are delineated by black triangles.


```

HnFSHR  MLLWVYHARQRTPTMHQVWMLL--LGLL-PVLLDOKLVEAHGCFNSRNTSCSFFCLGAKIHOMPSTIPWNTSSIEFRLTQLRTFPPKAFWG 87
CcFSHR  M-----TKRMVLLMMLCFSLGWL-MSHXEGMLVGSHFCSFNGSTCNFFCLGNSVHEMPKHIPENTTFVEIKLTEIRVFFRAALSE 79
OmFSHR  MM-----KMKKIMKMLLCLMGCVCVSOAEVAMVNS-----GTTFTYLCMGNITITHMPHIPKNTDLEFKQTHIRVFPREAFNTN 74
HnFSHR  LTEVSRIMLSENGLEEIGPQAFYNLSKLVETITKSKHLVVIHKDAFLDLPKLEYLTMNTGLKLLPDFSKINSAAHGFLLDLQDNMHI 177
CcFSHR  LHELKRIVVSENGALERIEPFAFSNLTELQEIITKSKNLVSL-KDAFWSLPKLRYLTSNTGLKALPDFSKINSAALEFLFDLQDNMHI 168
OmFSHR  LQQLTAVLTENGMLESIGAFANLPRLEITITKSKHLVVIHHCQAFGLPKLSHLTICNTGLRVLNFSRHSAAITFLDLQDNMHI 164
HnFSHR  DVIPPN AFLGLSADTIKELRLTKNGITEVRS HAFNGTKMOKLSLMGNQLRQIHRHAFMGAEGPIVLDISR TAVSTLPENMLRGIKLLTA 267
CcFSHR  DKIPPN AFLGLSATITELRLTKNGIRAIESYAFNGTIEKFLMGNQQLSHIDRYAFKGAEGPIVLDISHTAVHTLPENMLRGIKLLTA 258
OmFSHR  VIIPSN AFLGLTINTIDELRLTKNGISEVES HAFNGTKIHKFLMGNLQLSHMHNNSFKGAEGPGLDISRTALSLPE SVLGEVEHLSA 254
HnFSHR  VGVHNLKLPRLLELFTQIEANLTYPSHCCAFANSRKNMSVE-NSMCSLPHIOEDEPHFFLDHCWDVSKVSCFPKPD AFNPCEIIMGFTY 356
CcFSHR  TSVYSLKRLPNLELFTLQANLTYPSHCCAFKNFTKHKSVK-NQMCNNSGAPS-EPYFFEDHCKDVIKVTCTPPTDAFNPCEDIIMGFTF 346
OmFSHR  VSVESLALPPLSLFTKLRQANLTYPSHCCAFHKQRNRTFRMTSACFKPGAQN-NLHFFMDFCLNWTSVACSPAPDAFNPCEDI MGSAP 343
HnFSHR  LRVL IWIIVSVAIAGNLLVLA VLLSSHYKLTVPRFLMCHLAFADLCMGIYLLLIATVDTRTRSHYYNHGIEWQTGPGCATAGFFTVFASE 446
CcFSHR  LRVL IWIIVSILAIVGNCVLLVLFSSRYKLTVPRFLMCHLAFADLCMGIYLLIIAAKDITHTQSRYYNYGIDWQTGPGCHVAGFFTVFSE 436
OmFSHR  LRVL IWIIVSVLALLGNTIVLLVLLSRAKMTVPRFLMCHLSFADLCMGIYLVVVIATVDVTRTRGLYYNHAISWQTGAGCDIAGFFTVFASE 433
HnFSHR  LSVYTLTAVTLERWHTITFALRVDRLRLRHACAIMAVGWLFA SLAAVLPVVGSSYSKVSICLPMDVETGAAQVYVVVLLLLLNVAFLA 536
CcFSHR  LSVYTLTAVTLERWHTITYAMORERQMLRHACAIMAVGWLFA LLTALMPVGVSSYKKTISICLPMDVETVISQYVVVLLLLLNVAFLI 526
OmFSHR  LSVETLTAVTLERCHTITHALRLDRKLRRLRHACAVMATGWA FSCLAALLPTVGVSSYSKVSICLPMDVESLPSQVEVME LLLLLNVAFLC 523
HnFSHR  VEVCYLRIYLTVRNPFVFPASADM-RVAKRMAVLI FTDFLCMAPISFFAISAAKLP LITVSHAKVLLVLFYPI NSCSNPFLYAFFTKTF 625
CcFSHR  VCVCYMR IYLTVHNPTFVPANADM-RIAKRMAVLI FTDFLCMAPISFFAISAAKLP LITVSHAKVLLVLFYPI NSCSNPFLYAFFTKTF 615
OmFSHR  VCVCYLS IYLSVRNSSPPASAETS RMAQRMAI I FTDFLCMAPISFFALSAALKLP LITVSDSKLLLVLYPINS CANPFLYGLCTR TF 613
HnFSHR  KRDFFILASRFGCFKARAOMYRTEISSVONGGWP-THKTSDATL YSLV PNVNHVH 679
CcFSHR  KRDFFILTSQFGCFKTRAH IYRTEISSGONGAKVP-SPKTS DGTLYSLGHITQVY 669
OmFSHR  HRDFFLLAARYGLFITTKAQYRTEISFPVQDAAWIQMSPK VSHGTLCD----- 659

```

Figure 14c. Alignment of deduced protein sequences of *Heterotis niloticus* follicle-stimulating hormone receptor (HnFSHR) with FSHRs of *Oncorhynchus mykiss* FSHR (OmFSHR) and *Cyprinus carpio* FSHR (CcFSHR) showing little conserved domains in gray.

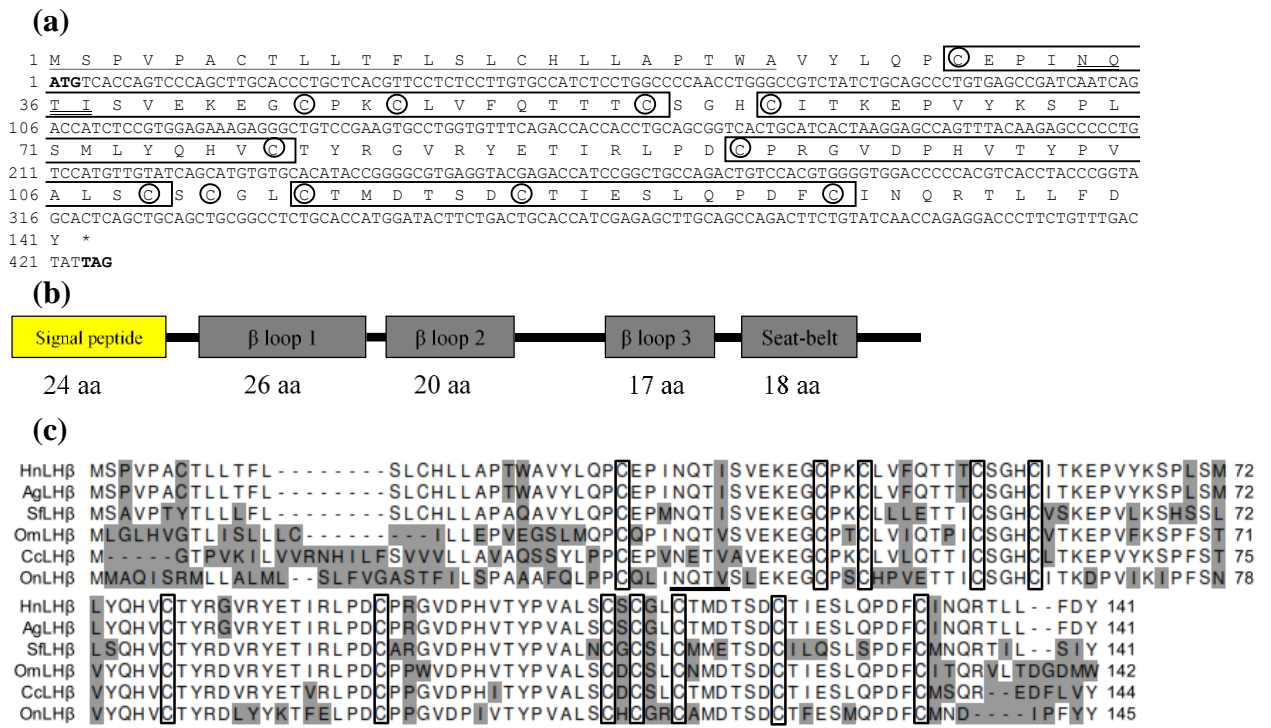


Figure 15. Nucleotide sequences and deduced protein sequences of (a) the beta subunit of *Heterotis niloticus* luteinizing hormone (HnLH β), box modeling of HnLH β with the different domains (b) and alignment (c) with *Arapaima gigas* LH β (AgLH β), *Scleropages formosus* LH β (SfLH β), *Oncorhynchus mykiss* LH β (OmLH β), *Oreochromis niloticus* LH β (OnLH β) and *Cyprinus carpio* LH β (CcLH β). In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, β -loops 1, 2 and 3 and the “seat-belt” are laid out successively in boxes. The predicted N-glycosylation site is double-underlined. In figure (c), conserved cysteines across species are in boxes, the little conserved domains are in gray.

(a)

```
1 M Q I L P V F L I L Q S I L F S L T S C H V F V C P G I C C C S S E T I
1 ATGCAGATTTGCCCCGTTCCTCATCTCCAATCGATTTGTTCAGTTTGACCTCATGCCACTTCGTGTGCCGGGGATTGCTGCTGTTCCTCCGAGACGATC
36 R C T E T T E R S I P G I Q I D A F K R L T V A H L S L S S I S S H A
106 CGCTGCACCAGTACAGAGGATCAATTTCCGGGATCCAGATAGATGCTTTCAAGCGACTGACTGTTGCTCACCTGTCCCTGAGCAGCATCAGCAGTCATGCC
71 F S G L S G V S R I E I T Q S D T I N T I E A M A F N N L I N I S E I
211 TTCACTGGACTCAGTGGAGTCTCAAGAATTGAAATTTACCAGAGTGACACCTTGAACACCATCGAGGCCATGGCCCTCAACACCTTCTAAACCTTTCTGAAATA
106 S I Q N T K N L V Y I N R H A F N K L P K L R Y L S I S N T G I T L I
316 TCAATACAGAACCAAGAATTGGTATACATCAATCGCCATGCATTCAACAAGCTGCCAAAGTTACGATATCTGAGCATTTC AACACTGGGATAACCTTGATC
141 P D L S S I S S L E S I F I L D I C D N L E L T T I P P N A F A G L S
421 CCAGATCTGTCAATCTTCTCTCGAATCCATTTTATCCTTGATATCTGTGACAACCTTCATCTTACCACCATTCCGCCAAATGCTTTTGTGGGCTATCA
176 N E Y A S M I L N G N G F K E V Q G Y A F N G T K I H K L I L S N N K
526 AACGAATACGCATCAATGATCTTAAATGGGAACGGTTTCAAGGAAGTTCAAGGTTATGCCTTCAACGGGACCAAGATTCAAACTGATCCTCAGCAACAAACAA
211 H L R I I H K D A F K G A I G F G I L D V S W T A L E M L P S Q G L Q
631 CATCTCAGAATAATCCACAAGATGCTTCAAGGGAGCCATTTGAGAGGATTTTAGATGTGCTGGACCGCCTGGAGATGCTCCCTCCAGGGATTTGACG
246 S L R L L V A H G A Y N L K T L P P L G A L G S I Q E A Q L T Y P S H
736 TCCCTCCGCTCCTGGTGGCCATGGGGCTTACAACCTTAAAGACCTCCACCCTGGGGCCCTAGGCAGCCTGACAGGAGCGCAGCTCACCTACCCAGTCAC
281 C C A L L A W N T G R D S S L F D D G N E S T H C K D S F S S E I P C
841 TGCTGTGCCCTGCTCGCCTGGAACACCGGCAGGGATTCTTCCTGTTTGTAGTGGGGAAACGAATCAACACTGTAAGGACAGCTTTTCATCGGAGATTCCAGGC
316 Q V S N S T V V G P L H E L M V F E E E L D T A D E V D F Q Y L D I D
946 CAAGTGTCTAATCCACTGTTGTTGGGCCACTGCATGAGTTAATGGTGTGGAGGAGCTGGACACTGCAGATCAGTGGACTTCCAGTACCTTGACATTGAC
351 F C Q N S R I L K C T P E A D A F N P C E D I A G F Q F L R V A I W F
1051 TTCTGTCAGAACTCCAGAACTGAAGTGCACCCAGAGGCAGATGCATTAACCCCTGTGAGGACATTGCTGCCTTCAATCTTGGGTTGCCATCTGGTTT
386 I N I L A I T G N I M V L L V L V A S R H K I S V P R F L M C H L A F
1156 ATCAACATACTCGCCATCACAGGTAACCTGATGGTGTCTACTGGTCTTGTGCGAGCCGCAACGCTGAGCGTCCCGCGCTTCCTCATGTGTACCTTGCCTT
421 A D L C I G L Y L L L I A I V D G R T Q G S Y S Q H A I A W Q T G P G
1261 GCCGACCTCTGATTTGGCCTTACCTCCTATTGATTTGCTATTGTTGGATGGGGCAGCAGGGCAGCTACAGCCAGCATGCCATCGCATGGCAGACAGGGCCAGGC
456 C E A A G F L S V F G G E L S V Y T I T A I T L E R W H T I S H A L Q
1366 TGTGAAGCTGCAGGCTTCCTGTCCGTGTTGGAGGTGAGCTGTCAGTGTACACATTGACCGCATCACTCTCGAGCGTGGCACACCATCAGTACGCCCCGCAA
491 P E R R I G I R E A L L A M A G V W L L C L G V A L L P L V G V S S Y
1471 CCGAARCGCGCCTGGGCTGAGACATGCTCTGCTTGCATGGCCGGCGTCTGGCTGCTGCCTGGGAGTGGCATGCTCCCTCGTGGGTGTCAGCAGTTAT
526 R K V S V C L P M D I D T P L A Q A F V I L L L L L N V T A F L V V C
1576 AGGAAGTGTAGTGTGCTGCCATGGACATCGACACACCCCTGGCCAGGCTTTTGTATTCTCCTCTCTCCTCAATGTAAGTGCCTTCTGGTCTGCTGTG
561 G C Y V R I Y Q A V R N P K F A G R S A D A R I A K R M A V L I F T D
1681 GGCTGCTATGTGCGGATCTACCAGGCAGTGGCAACCCCAAGTTTGGCGGAAGAAGTGGGATGCCAGGATTGCTAAGCGCATGGCAGTGTCTCATCTTCACTGAT
596 F L C M A P I S F F A I S A A F R M P L I T V T N S K I L L V L F Y F
1786 TTCTCTGCATGGCACCACTCTCCTTTTTCGCTATCTCTGCTGCCTCAGAATGCCTCAATTAATCTGTGACCACTCGAAGATCCTGCTGGTCTCTCTATCCC
631 I N S C A N P F F L Y A I C T K A F R R D V F A L A S T L A C C E S K A
1891 ATCAACTCCTGTGCCAACCCCTCCTCTATGCCATCTGTACCAAGGCGTTCGGAGAGATGTGTTTGGCCTTGGCCAGTACGCTGGCCTGCTGCCAGAGCAAGGCC
666 S V Y R T K A Y C L D N S G D K T S G S V A A K K S S H V T L K T A A
1996 AGTGTGATCGCACAAAGGCTACTGCCTGGACAACAGCGGTGACAAGACGAGTGGCTCAGTAGCTGCCAAGAAGAGCTGCACGTGACACTCAAAACAGCTGCC
701 L S A S I P V E *
2101 CTCAGTGCATCATTACCAGTAGAGTAA
```

(b)

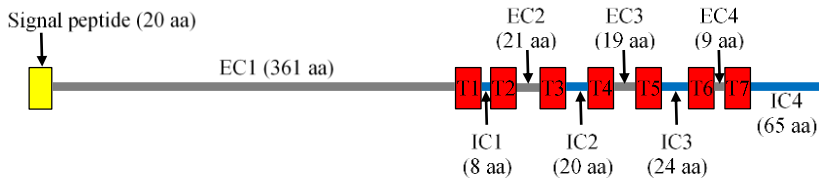


Figure 16. Nucleotide sequences and deduced protein sequences (a) of *Heterotis niloticus* luteinizing hormone receptor (HnLHR), and box modeling (b) with the different extracellular domains (EC1 to EC4), transmembrane domains (T1 to T7) and intracellular domains (IC1 to IC4) of HnLHR. In figure (a), the start codon and the stop codon are in bold type, the signal peptide is underlined, T domains are in boxes, IC domains are double-underlined, and EC domains are delineated by black triangles.

N-ter1 HnVtg1 (Blood)

MRAVLLALTLALAAGQDLSLTPDFANGKAYMFKYEAE LLGLPVEGLGKAGVKIVCKVLI SRVSQNTYLLK LKEPQIFEYSGIRPRDDFKPAAKLTQALAAQLLIPV
 KFEYIRGVVGVKVFAPA AVSKTILNLRHGI LNILHVNIKKTQNVYELQEAGSQGVCKTDYAI SEDTKAEHIYVTKSKDLGNCQNRIMADIGIAYTETCVQCQMRNKNL
 RGAATYSYVMKPTETGALITKATVQEVHQFTPFHELRLGAGQVETRQEMTFVETQNDPVQPAHADYVARGSLQYEFASELLQTPLEPKITDIKTQIEEILNHLVKN
 MGEVHEDAPLKFLQTLQLLREAKYEMNGIWTQVKSQKPVFRWFVLDTPVAVGNQDAVRFI KEKFIAGDISAAETAQALLVALHLEANMDTVNLAGTLVHFAKLQSH
 PMLREIAMLGYGSIVFVFKLCTKQONCPAEVTKPVHDIAAEAISKANVEEIALVLKVLGNAGHPASIKPIMKVLPFGFSTASSLTVKVHVDAVMALRHI AKRENHVKVQE
 VALQLFMNRDLHPEVRM VACVVFLECKPPIGLVAMIASALQNEKSLQVASFAYSHMKALTRSTAPELAQVAACNVAIKILSPKLDRLSRYFSKTIHMDFFNYHLMA
 GAAATAHLINDAASILPRALVAKIRACMLGAAADVLEFGVTEGLQEA LLKSPAADPNADRLTRMWRV LNTLKNWKS LPASQPVASAYLKWFGQEIAFANIDRDI IE
 KAVEFATGAAAQPALLKNILNMMQSGIDIQIAKPLMTSEVRRIFFPTSMGFPIEISLYSAAVA AAIKAKAAILNPKPSDNFRI AQLMNTDIQLDTRVVP SIAVHKYAV
 MGVNTALIQAAIEAKVKVQKVLPLKFNARINIAQGHYKIE TVPLHAQERILDLHMETVAVARNIENLSEAKITPVLPARLASQSKETFASAGSGSKSSERIEHES
 SSHPMQHAVSARTDRQWCVTVASLADQACAKVTSQNA GFIRNSPLYKLI GEHSVIAAVKVPVSGEAVDKIEIEMHVGPDAASKIVKTI TVKDDNAKEGHAGHSV PVLK
 LREILETEKKQHSRNATTSSSSSSSSRRSGKSSSSSSSSSSSSSSSSSSSSSSSSS SSKRSKKNIKR SSSSSSSSSSSSSSSSSSSSSSSRRRTENVLGGSSSSSSSSRSRSQTSKAAIF
 QRFTQNHQHETTRAASSQKTSSSPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS

N-ter1 HnVtg1 (Mucus)

MRAVLLALTLALAAGQDLSLTPDFANGKAYMFKYEAE LLGLPVEGLGKAGVKIVCKVLI SRVSQNTYLLK LKEPQIFEYSGIRPRDDFKPAAKLTQALAAQLLIPV
 KFEYIRGVVGVKVFAPA AVSKTILNLRHGI LNILHVNIKKTQNVYELQEAGSQGVCKTDYAI SEDTKAEHIYVTKSKDLGNCQNRIMADIGIAYTETCVQCQMRNKNL
 RGAATYSYVMKPTETGALITKATVQEVHQFTPFHELRLGAGQVETRQEMTFVETQNDPVQPAHADYVARGSLQYEFASELLQTPLEPKITDIKTQIEEILNHLVKN
 MGEVHEDAPLKFLQTLQLLREAKYEMNGIWTQVKSQKPVFRWFVLDTPVAVGNQDAVRFI KEKFIAGDISAAETAQALLVALHLEANMDTVNLAGTLVHFAKLQSH
 PMLREIAMLGYGSIVFVFKLCTKQONCPAEVTKPVHDIAAEAISKANVEEIALVLKVLGNAGHPASIKPIMKVLPFGFSTASSLTVKVHVDAVMALRHI AKRENHVKVQE
 VALQLFMNRDLHPEVRM VACVVFLECKPPIGLVAMIASALQNEKSLQVASFAYSHMKALTRSTAPELAQVAACNVAIKILSPKLDRLSRYFSKTIHMDFFNYHLMA
 GAAATAHLINDAASILPRALVAKIRACMLGAAADVLEFGVTEGLQEA LLKSPAADPNADRLTRMWRV LNTLKNWKS LPASQPVASAYLKWFGQEIAFANIDRDI IE
 KAVEFATGAAAQPALLKNILNMMQSGIDIQIAKPLMTSEVRRIFFPTSMGFPIEISLYSAAVA AAIKAKAAILNPKPSDNFRI AQLMNTDIQLDTRVVP SIAVHKYAV
 MGVNTALIQAAIEAKVKVQKVLPLKFNARINIAQGHYKIE TVPLHAQERILDLHMETVAVARNIENLSEAKITPVLPARLASQSKETFASAGSGSKSSERIEHES
 SSHPMQHAVSARTDRQWCVTVASLADQACAKVTSQNA GFIRNSPLYKLI GEHSVIAAVKVPVSGEAVDKIEIEMHVGPDAASKIVKTI TVKDDNAKEGHAGHSV PVLK
 LREILETEKKQHSRNATTSSSSSSSSRRSGKSSSSSSSSSSSSSSSSSSSSSSSSS SSKRSKKNIKR SSSSSSSSSSSSSSSSSSSSSSSRRRTENVLGGSSSSSSSSRSRSQTSKAAIF
 QRFTQNHQHETTRAASSQKTSSSPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS

C-ter HnVtg1 (Blood)

SSSSSSSSSSSSSSSSSSGSSASSFENIYKKSRLFGD TVRPAAVVIVRAVRGNERRRGYQVAAYMDKANARVQVI ISALADSDKWQLCADGIQLSMHKVMKIRW
 GAECQDYRAVIKAETG LLGPHPAAQ LKMYWNKTPRALKRYASMIYEYVPGVALLAGFSEGRHRSGERQIKL TMAATSARTISI ILRTPWMTLYKLGVIPIALPIGA
 AAAARAEVQT FAGRIHYMFVEATS AKCKLENSTVTTFNNRRYGLQMPRSCYQVVAQDCTSKLKF MVLRKGDERSSEESHVIVKIADIDVDLTAEHGNIQVKVNGRVVP
 ITQHNYEHPTGTISIKQK GAGISLCAPSHGLHEVYFNKNILTIQVPDWMKG VCGLCGKADG D VRRFQQGPSGHIE DAVSFAHSWVLA AESCHDAKQCVKQELFK
 LTEPVLNDQDMKQSTFPVLSCLPQCS PMKTPVTVGFHCIPIDSNVNSWSS IRKKQEDIRVTVDAHIMCNCGEECA

C-ter HnVtg1 (Mucus)

SSSSSSSSSSSSSSSSSSGSSASSFENIYKKSRLFGD TVRPAAVVIVRAVRGNERRRGYQVAAYMDKANARVQVI ISALADSDKWQLCADGIQLSMHKVMKIRW
 GAECQDYRAVIKAETG LLGPHPAAQ LKMYWNKTPRALKRYASMIYEYVPGVALLAGFSEGRHRSGERQIKL TMAATSARTISI ILRTPWMTLYKLGVIPIALPIGA
 AAAARAEVQT FAGRIHYMFVEATS AKCKLENSTVTTFNNRRYGLQMPRSCYQVVAQDCTSKLKF MVLRKGDERSSEESHVIVKIADIDVDLTAEHGNIQVKVNGRVVP
 ITQHNYEHPTGTISIKQK GAGISLCAPSHGLHEVYFNKNILTIQVPDWMKG VCGLCGKADG D VRRFQQGPSGHIE DAVSFAHSWVLA AESCHDAKQCVKQELFK
 LTEPVLNDQDMKQSTFPVLSCLPQCS PMKTPVTVGFHCIPIDSNVNSWSS IRKKQEDIRVTVDAHIMCNCGEECA

(a)

HnVtg1 MRAVLLALTLALAAGQDLSLTPDFANGKAYMFKYEAE LLGLPVEGLGKAGVKIVCKVLI SRVSQNTYLLK LK 73
 HnVtg2 MKAAVVALALAVAGQQNILTPDFAAKKTVYKYEAQLHSELPEEGLAKAGLK IASKVLI SRADQNTYLLK LK 73
 * * ** *** **** ***** * * ** ** * ** ** * * ** * * ** * * ** *

(b)

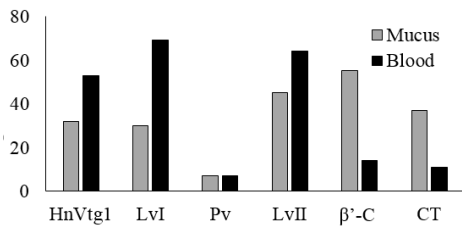


Figure 17. Coverage of the HnVtg1 N-ter and C-ter peptides detected in *Heterotis niloticus* blood and mucus by mass spectrometry. (a) Peptides specific to HnVtg1 and HnVtg2 detected in the blood. The amino acids conserved between the two Vtg are indicated by stars. Detected peptides are in gray, the signal peptide is underlined. (b) Coverage rate (in percentage) of the HnVtg1 peptides detected in the blood and mucus compared to the whole protein (N-ter1+C-ter) and the different domains. Only Vtg1 was detected in mucus.