



Article

Gene Characterization of Nocturnin Paralogues in Goldfish: Full Coding Sequences, Structure, Phylogeny and Tissue Expression

Diego Madera, Aitana Alonso-Gómez , María Jesús Delgado , Ana Isabel Valenciano and Ángel Luis Alonso-Gómez *

Departamento de Genética, Fisiología y Microbiología, Universidad Complutense de Madrid, 28040 Madrid, Spain; dmadera@ucm.es (D.M.); aitaalon@ucm.es (A.A.-G.); mjdelgad@ucm.es (M.J.D.); aivalenc@ucm.es (A.I.V.)

* Correspondence: alalonso@ucm.es

Abstract: The aim of this work is the full characterization of all the nocturnin (*noc*) paralogues expressed in a teleost, the goldfish. An in silico analysis of the evolutive origin of *noc* in Osteichthyes is performed, including the splicing variants and new paralogues appearing after teleostean 3R genomic duplication and the cyprinine 4Rc. After sequencing the full-length mRNA of goldfish, we obtained two isoforms for *noc-a* (*noc-aa* and *noc-ab*) with two splice variants (I and II), and only one for *noc-b* (*noc-bb*) with two transcripts (II and III). Using the splicing variant II, the prediction of the secondary and tertiary structures renders a well-conserved 3D distribution of four α -helices and nine β -sheets in the three *noc* isoforms. A synteny analysis based on the localization of *noc* genes in the patrilineal or matrilineal subgenomes and a phylogenetic tree of protein sequences were accomplished to establish a classification and a long-lasting nomenclature of *noc* in goldfish, and valid to be extrapolated to allotetraploid Cyprininae. Finally, both goldfish and zebrafish showed a broad tissue expression of all the *noc* paralogues. Moreover, the enriched expression of specific paralogues in some tissues argues in favour of neo- or subfunctionalization.



Citation: Madera, D.; Alonso-Gómez, A.; Delgado, M.J.; Valenciano, A.I.; Alonso-Gómez, Á.L. Gene Characterization of Nocturnin Paralogues in Goldfish: Full Coding Sequences, Structure, Phylogeny and Tissue Expression. *Int. J. Mol. Sci.* **2024**, *25*, 54. <https://doi.org/10.3390/ijms25010054>

Academic Editor: Kiyoshi Naruse

Received: 30 October 2023

Revised: 5 December 2023

Accepted: 14 December 2023

Published: 19 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: nocturnin nomenclature; *Carassius auratus*; zebrafish; molecular cloning; gene evolution; synteny analysis; synteny quantification; phylogenetic tree; protein structure; tissue distribution

1. Introduction

Nocturnin (NOC) is an enzyme discovered in the retinal photoreceptors of *Xenopus laevis*, where it displays a daily rhythm of expression peaking at the beginning of the scotophase [1]. Day–night variations in the levels of *nocturnin* (*noc*) gene expression have been described in species belonging to different animal phyla. In Porifera, the *noc* expression is found in the demosponge *Suberites domuncula* in the epithelial layer around the aquiferous canals, showing maximal levels in darkness [2]. In insects, a daily rhythm of *noc* expression is described in the brain of female mosquitos [3], and in both the head and trunk of the fruit fly (*Drosophila melanogaster*) [4]. In vertebrates, a ubiquitous pattern of *noc* expression is also present from fish [5,6] to mammals [7,8], including the central nervous system (retina and encephalic areas) and most of the organs. In amphibians, *noc* was detected by in situ hybridization in most of the tissues during the early development of clawed frog (*Xenopus laevis*) [9]. In mammals, a daily variation of mouse *noc* expression has been also reported in the retina and brain areas, and with higher amplitude in non-neural tissues [7]. Besides, a daily variation of *noc* expression has also been detected in human peripheral blood lymphocytes [10].

The nuclear regulation of *noc* expression has been only studied in a few species of vertebrates. In *X. laevis* photoreceptors, the binding of pCREB (phosphorylated cAMP response element-binding protein) to a sequence in the nocturnin promoter activates the

transcription of *noc*. Light inhibits CREB phosphorylation, while darkness induces pCREB and, therefore, the transcription of *noc* gene [11]. In a hepatoma cell line, *noc* expression is controlled by the binding of the CLOCK/BMAL1 heterodimer to an E-box element of the *noc* promoter, similarly to the activation of other clock-controlled genes [8].

Possible functions of NOC remain quite elusive to date. Due to its structural homology with the *ccr4* (carbon catabolite repression 4) family, NOC was first linked to the regulation of the mRNA half-life in cells, removing the poly-A tail of some clock genes and clock-controlled genes [12,13], and therefore with a possible role in the circadian system [14]. Further studies demonstrated that the main enzymatic activity of NOC is the dephosphorylation of NADP(H) to NAD(H) [15], linking NOC with the metabolic homeostasis as a switch between catabolism and anabolism. Accordingly, some data relate *noc* expression with the metabolic control of energy balance, and this possible relationship seems to be well-conserved throughout phylogeny. Thus, in demosponges, the daily rhythm of *noc* expression is inverted with respect to glycogenin [2]. In fruit flies, *noc* expression is upregulated after starvation [4]. In fish, feeding regulates *noc* expression in the hypothalamus, hepatopancreas, and intestinal bulb [6]. In mammals, knockout mice (*noc*−/−) showed resistance to obesity and hepatic steatosis when they were fed a high-fat diet [16]. Furthermore, the *noc*−/− mice have also deficits in the intestinal triglyceride absorption [17] and dysregulation of lipids and glucose metabolism [18,19].

In addition to metabolic functions, NOC is also involved in other functions; for example, it may act as a non-enzymatic transcription co-factor by promoting the transcription factor PPAR- γ to translocate into the nucleus of fat cells [20], enhancing the differentiation of preadipocytes to adipocytes. Moreover, NOC negatively regulates osteogenesis by inhibiting bone IGF-1 expression [14,21], and *noc* is critical for the development of early mouse embryo [22], in the timing and differentiation of somites in amphibians [23], and in wing morphogenesis at the pupal stage in fruit flies [4].

A full-length cDNA sequence of *noc* has been cloned to date only from the retina of the clawed frog [1], mouse [7], and one of the two *noc* paralogues (*noc-b*) in zebrafish [5]. Besides, partial cDNAs from humans, cows, chickens, zebrafish, and goldfish have also been obtained [5–7]. Nowadays, thanks to massive sequencing technology, the entire genomes of many vertebrates are now available. The in silico analysis of these data allows us to conclude that all vertebrates so far studied have at least one gene coding for *noc*, with the same basic structure consisting in three exons with a highly conserved sequence.

Teleosts are a very interesting vertebrate group used to address genomic studies, because of the whole-genome duplication (named 3R) undergone by their ancestor [24]. This fact determines that current teleosts should have at least two gene paralogues for *noc*. In gene sequence databases, these two genes were firstly called “nocturnin” and “nocturnin-like”, respectively, being later renamed *nocturnin-a* and *-b* (*noc-a* and *noc-b*), based on full and partial sequences obtained from cDNA in zebrafish [5] and goldfish [6]. Particularly, both salmonids [25] and the subfamily Cyprininae (to which goldfish belongs) [26] go through an additional whole-genome duplication, known as 4Rs and 4Rc, respectively. Thus, the presence of four *noc* paralogues in the genome of goldfish is expected.

The aim of this work is to carry out the first full characterization of all *noc* paralogues expressed in the Cyprininae subfamily. We have obtained the complete coding sequences of all splicing variants of *noc* paralogues expressed in the goldfish transcriptome, as well as the mRNA non-coding regions (5'UTR and 3'UTR). A comprehensive in silico analysis of the evolutionary profile of these genes has been performed, including the synteny, phylogenetic tree, and prediction of secondary and tertiary protein structures deduced from translated amino acid sequences. Furthermore, we have studied the tissue expression of *noc* transcripts in two species of cyprinids, goldfish (*Carassius auratus*) and zebrafish (*Danio rerio*). This full characterization of all the *noc* paralogues in the Cyprininae subfamily lays the essential groundwork for further functional studies of this intriguing enzyme.

2. Results

2.1. Teleostean Nocturnins: Evolutionary Origin

The general structure of the gene encoding *noc* has not changed throughout the teleost phylogeny (Figure 1). Taking the spotted gar (*Lepisosteus oculatus*), a non-teleost model as ancestral Osteichthyes prior to the 3R duplication, we found that the single *noc* gene has three exons and two introns (Gene ID: 102682701). A detailed search in the genomic sequences of this species allows us to propose the existence of three alternative sequences for the first transcribed exon. We named these exon 1a, exon 1b (non-coding), and exon 1c. By means of alternative splicing, three different transcripts for *noc* may be generated. Based on the position of transcription start sites (TSS) in the gene sequence, we named them variant I (with exon 1a, accession no. XM_006629623.2), variant II (with exon 1b, accession no. XM_015345017.1), and variant III (with exon 1c, accession no. XM_015345016.1). To date, no experimental confirmation of these variants on transcriptomes has been published.

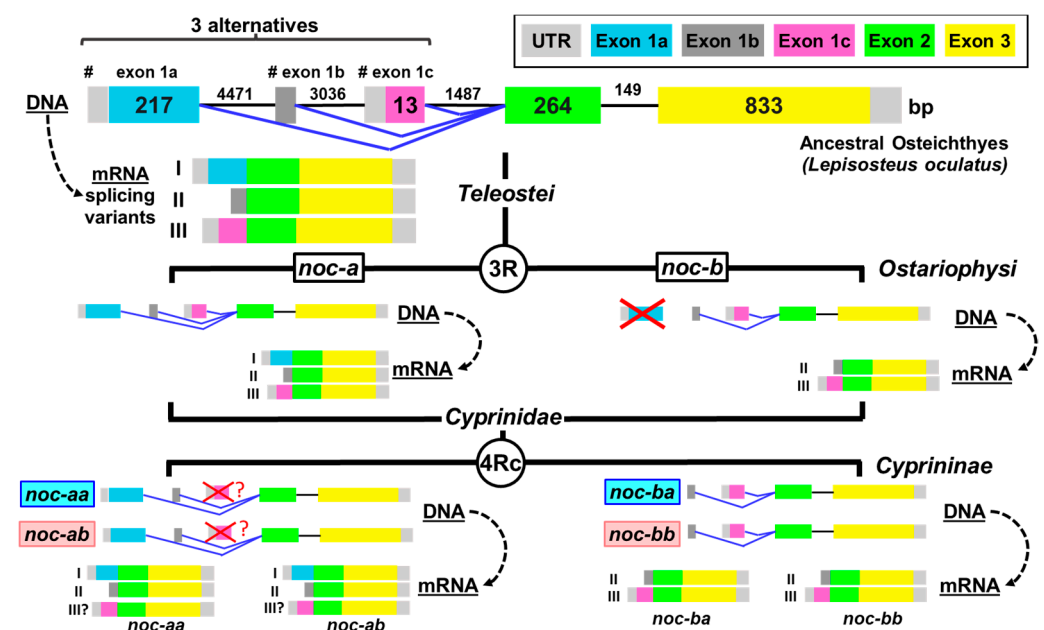


Figure 1. Nocturnin gene structure in Osteichthyes. Full-genome duplications are indicated as 3R (Teleostei specific) and 4Rc (Cyprininae specific). Proposed model for the exon–intron structure of the gene encoding nocturnin in different Osteichthyes clades, and its transcription into mRNA (dashed lines). The I, II and III indicate the alternative splicing variants. # indicates the alternative transcription start sites. Exons are represented by boxes and introns by lines. Angled blue lines indicate alternative splicing of the first exon (exon 1a, exon 1b or exon 1c) in mature mRNA. The length (bases pair, bp) of exons (inside the boxes) and introns (above lines) is indicated for the *Lepisosteus oculatus* gene. Lost exons during evolution of Ostariophysii and Cyprininae are represented by red crosses. Question marks indicate unconfirmed loss of variant III in the *noc-a* paralogues of cyprinines. Light pink and cyan backgrounds on Cyprininae nocturnin names indicate matrilineal and patrilineal origin, respectively.

The 3R genomic duplication in the common ancestor of teleosts (350 million years ago) resulted in two forms of *noc* (*noc-a* and *noc-b*) that were initially identical but have accumulated several divergences (by sub-functionalization and/or neo-functionalization) during their subsequent evolution, which allowed their conservation in current teleosts. The analysis of the sequences shows that *noc-a* seems to retain the three alternatives for exon 1 in many species, but the *noc-b* gene loses the exon 1a (splice variant I) in the ancestral Clupeocephala (advanced teleosts: Otomorpha plus Euteleostei), since the basal teleostean fish *Anguilla* sp. (accession no. XM_035419224.1) and *Scleropages* sp. (accession

no. XM_018737923.2) do conserve exon 1a. Then, *noc-b* for Clupeocephala only has two splicing alternatives (forms II and III, Figure 1).

Furthermore, the *noc-b* gene structure underwent additional changes in teleostean fish, a new short intron (phase 0) splits exon 3 into two parts in the Euteleostei clade (e.g., *Salmo salar*, Gene ID: 106610553), and other phase-0 intron splits exon 2 in the Neoteleostei clade (e.g., clown anemonefish, *Amphiprion ocellaris* Gene ID: 111585149); however, these insertions did not break the coding sequence of *noc-b*. In the Ostariophysi clade, these new intron insertions were not observed, preserving the ancestral structure. In orders Characiformes, Siluriformes, and Gymnotiformes, the *noc-b* gene has been entirely lost.

The four paralogues for *noc* that would be expected after the additional duplication 4Rc of the Cyprininae subfamily are confirmed in genomic databases of *Cyprinus* sp. and *Sinocyclocheilus* sp. The 4Rc duplication generates allotetraploids, where the karyotype is divided into two subgenomes with a female or male origin. Then, we named these *noc-aa*, *noc-ab*, *noc-ba*, and *noc-bb*, where the last letter indicates patrilineal (a) or matrilineal (b) subgenomes. The cyprinine *noc-a* paralogues (*noc-aa* and *noc-ab*) have been annotated in databases only with two splicing variants, I and II, suggesting the loss of variant III (Figure 1).

After sequencing the *noc* mRNA of goldfish (Figure 2), we obtained evidence of the two isoforms for *noc-a* (*noc-aa* and *noc-ab*), with only two splice variants (I and II). For *noc-b*, only one paralogue (*noc-bb*) has been found with two possible transcripts (II and III). Despite not having found any trace of *noc-ba* in our goldfish mRNA samples, it is possible that *noc-ba* could exist in the goldfish genome, as it is found in other members of the Cyprininae subfamily that also underwent the 4Rc duplication. Therefore, we extend our search to genome projects.

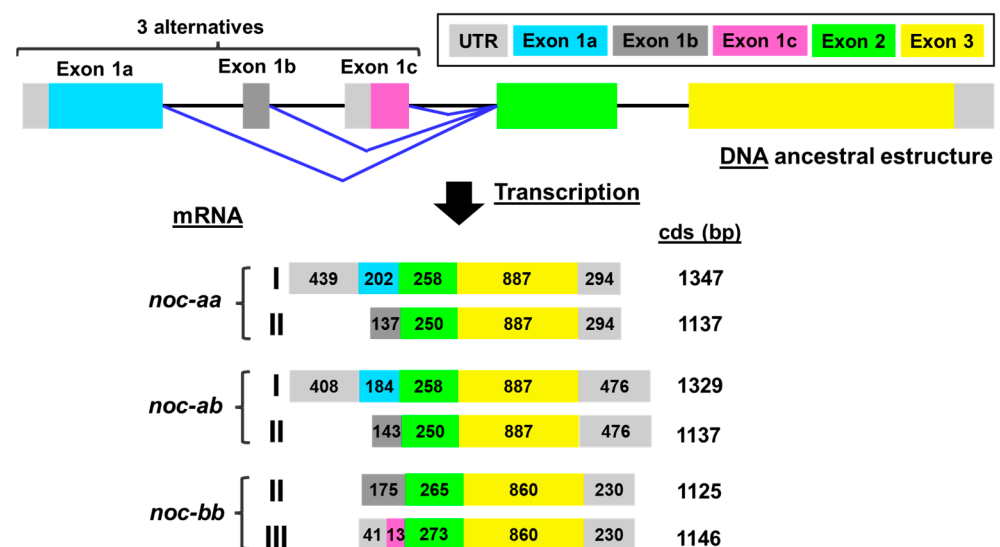


Figure 2. Exon-intron structure and transcription pattern of *nocturnin* gene paralogues in goldfish obtained by mRNA sequencing. I, II, and III indicate the alternative splicing variants. Exon 1a, exon 1b, and exon 1c are the alternatives of the first exon in mature mRNA. Angled blue lines indicate an alternative splicing mechanism. Coding exons are indicated by colored boxes and introns by lines. Grey boxes indicate untranslated regions (UTR and exon 1b). The length (bases pair, bp) of exons is indicated inside the boxes. The length of coding sequences (cds) is also indicated.

Accessing the complete genome of *Carassius gibelio* (Prussian carp) and *Carassius auratus*, we found the paralogue *noc-ba* with a high identity with the *noc-ba* sequences of other cyprinines (such as different species of *Sinocyclocheilus* and *Cyprinus carpio*). However, the *Carassius* sp. sequences show several mutations, especially large deletions in the coding part, resulting in truncated and non-functional coding sequences (Figure 3). Some of these deletions are identical in *C. auratus* and *C. gibelio*. Such deletions are confirmed to be real,

and are not artifacts from the sequence assembly process. The consequence of frameshift deletions is the appearance of premature stop codons in exon 3 (13 red stars in *C. gibelio*, and 3 in *C. auratus*, Figure 3).

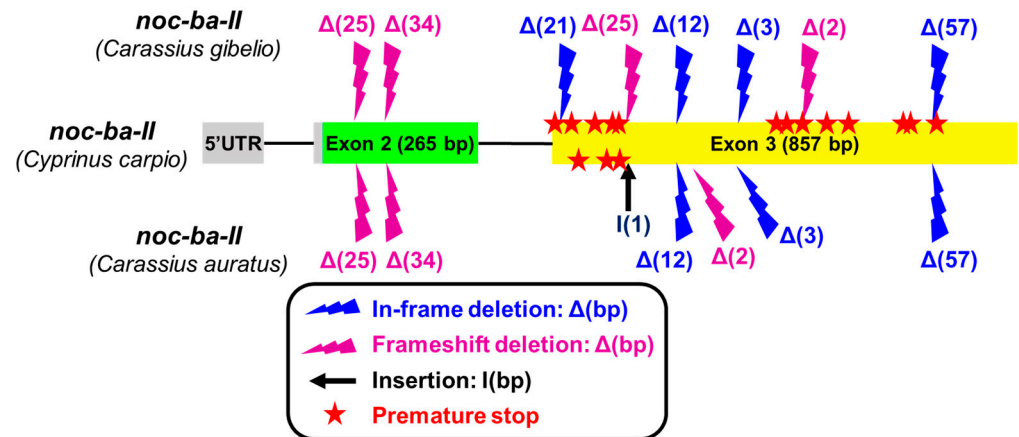


Figure 3. Mutations accumulated in the *noc-ba* paralogue of *Carassius gibelio* (Gene ID: 127988217) and *Carassius auratus* (Gene ID: 113105808) compared to *Cyprinus carpio* (accession no. XM_019094354.2). Illustration represents the exon–intron structure of the splicing variant II of *noc-ba*. Exons are indicated by boxes and introns by lines. The symbols above (*C. gibelio*) and below (*C. auratus*) indicate the mutations compared to *C. carpio* sequence. The blue and purple rays represent in-frame and frameshift deletions, respectively, and the base pairs suppressed, $\Delta(\text{bp})$. The black arrow I(1) indicates an insertion of 1 bp. The red stars represent premature stop codons in transcribed sequence.

2.2. Nocturnin Sequences in Goldfish

2.2.1. Transcription of Nocturnin aa (*noc-aa*) and Nocturnin ab (*noc-ab*) Genes

Sequencing of mRNA allows for the description of complete coding sequences for *noc-aa* splicing variants I and II (Figure 4). As expected from proposed cyprinine *noc* evolution (Figure 1), no trace of variant III was found. Exons 2 and 3 are common to both variants (I and II); however, the exon 1 is different. For variant I, there is a long coding sequence (202 bp) from initial methionine (exon 1a). Variant II has a different exon 1b, with no available methionine codon (ATG). Instead, the first codon ATG in exon 2 initiates variant II (green box in Figure 4).

By cloning the RT-PCR amplicons, we resolved the multiple single-nucleotide polymorphisms (SNPs) that we found using conventional PCR. We have identified two alleles of the *noc-aa* paralogue, with two splicing variants I and II, respectively (Figure 4). The differences between both allele sequences are in the 5' UTR and in the exons 1 and 2, showing 11 changes in nucleotides. Despite this, we found only four amino acid conservative changes (Val/Leu⁸², Thr/Ala⁸⁹, Glu/Asp⁹⁷, Tyr/His¹⁰¹) that did not affect the structure of protein and were located far from the catalytic site.

As described for *noc-aa*, two transcripts are sequenced for the *noc-ab* gene (Figure 5), corresponding to the predicted splicing variants I and II. Variant II is not annotated in any available *Carassius* genome (NCBI and ENSEMBL databases). The short length of this non-coding sequence (exon 1b) may hinder detection during the automatized annotation process. However, variant II was recovered by our experimental mRNA sequencing or by searching in transcriptomic libraries (Supplementary Table S1). In this case, we did not detect alleles for *noc-ab*.

```

(exon 1b)
gac act gga tca aca ctg gat tag tcc tga agt ccc tca aag ttt gtg act tta gtt cac 29
ctg cat cag aga cat aca gtg gga aag gat ata ata cga g↓ 89
129

(exon 1a)
gcac tgg atg tcc ttt cgt tga cag aag gaa taa acc att caa tgt cct gta ggt gaa tat 61
atc act aaa ata tat tgc tat ttt gta ttt ttt cca tat acg ttt tgg att tgt ttg aaa 121
tgt tgg cat tcc aaa tgt agt ctt aat atg tca tat ata gaa taa ctt act act tta tgg 181
aat ttt att agt gca taa aaa cct aat cta ctg aac ccc cct atc aag ata tag tac acg 241
tgt cac cct gga aaa tga tga cac tta tac cgg ctt ttt tcc ata tag aca acg aaa cat 301
gtc taa gta gtg ccc aaa aat ctt tta caa atg tta tat agg ttt gta aac tta ttt ctt 361
cgt tag cca tac tta aag cca tag ttt ggg tta tca gtt taa act cgg gtt cga agg gct 421
gtc aac cga aga aga aca atg tat cgg acc agg cgc tgt tca tgg ctg ctt cac cga gaa 481
cca (allelic) M Y P T R R C S S L L H R E 14

cta gcc gca gtc tgt cta tca tca ctg ggg aca cat cca aag aga cac ggt cgg ctg aag 541
L A A V C L S S L G T H P K R H G P L K 34
aac agc agc ctt cgg agt tca cga ctc ttc cag acc agc gta att cat cct cct cca acc 601
N S S L P S S R L F Q T S V I H P P P T 54
agg tcc aac tca tca ccc ctc ctg ctc ggt tct tta caa gtg tgt cag atg ggt agc agc 661
R S N S S P L L L G S L Q V C Q M G S S (allelic)
agc agt aga ctc ttc agt act gtg gcc cag act ctg agc agt gct cgg ctg aca gac cca 721
S S R L F S T V A Q T L S S A A L T D P 94
cac gtg gat gct gac gac cat gag cat gag cag gct gac cct gac gct ctg ctg cgt gaa 781
H V D A D D H E H E Q A D P D A L L R E 114
tgt gag gag gtt tta agg aac cgg cct cct cga atg cac cga gac ttt gtt agg acc aga 841
C E E V L R N R P P R M H R D F V R T R 134
gcc tgc tca ctg cag aac aca ccc atc cgc atc atg cag tgg aac atc tta gca caa gct 901
A C S L Q N T P I R I M Q W N I L A Q A 154
ctg gcc gaa gcc aag gac ggt ttt gtg cgc tgt cgg atg gag gcg ttg aac tgg tca gag 961
L G E G K D G F V R C P M E A L N W S E 174
agg aag tat ctt ata ctg gaa gag atc ctt acc tac agg cca gat gta ctg tgt cta cag 1021
R K Y L I L E E I L T Y R P D V L C G L Q 194
gaa gtg gac cat tat ttt gac acc ttc cag cct gtg ctc gcc agc ttg gcc tac cag agc 1081
E V D H Y F D T F Q P V L A S L G Y Q S 214
agc ttc tgc ccc aag cgg tgc tcc ccc tgc cta gac gtt cac aat aac aat gcc cct gat 1141
S F C P K P C S P C L D V H N N N G P D 234
ggc tgc gct ctg ttt ttc aac cga cgt cgc ttc cag ctg ctc cac acc act cac ttt aga 1201
G C A L F F N R R R F Q L L H T T H L R 254
ctc tct gtc atg atg ctc aaa acc aac caa gtg gcc gtc gtg gct aca ctt cgc tgc agg 1261
L S V M M L K T N Q V A V V A T L R C R 274
ctc aca ggt caa gtc ttc tgt gtg gcg gtc aca cac ctg aaa gca cgg agc gcc gwg gag 1321
L T G Q V F C V A V T H L K A R S G G W E 294
tcg ttt cgt agc gca cag ggg gct cat ctc ctc cag cag ctc tgc gac cta acc aca cag 1381
S F R S A Q G A H L L Q Q L C D L T T Q 314
tca cac aca gag atg gat gat gag aac cgg acc agg gtt tgg gaa agc gaa gag ggt att 1441
S H T E M D D E N R T R V W E S E E G I 334
cgg ttg gtg gta tgt ggg gat ttt aat gcg gaa cca agt gag gaa gtg tac agg cgc ttt 1501
P L V V C G D F N A E P S E E V Y R R F 354
ctg tct tct ctc ctg ggg cta gac agc gct tat aag tgt ctg agt gag gac ggg aac aca 1561
L S S S L G L D S A Y K C L S E D G N T 374
gag cgg cct tac acc agc tgg aag atc cgg ccc agc gga gag agc tgc tct act tta gac 1621
E P P Y T S W K I R P S G E S C S T L D 394
tac att tgg tac tgg gaa aga gct ttc cag gtg gat gct gtg ctg agg ata cgg aac gaa 1681
Y I W Y S E R A F Q V D A V L R I P N E 414
gag cag atc ggg cca gac cgg ctt cgg tgg tat cat tat ccc tct gat cat ttg tct ctg 1741
E Q I G P D R L P S Y H Y P S D H L S L 434
gtg tgt gac tta agc ttt agt cag cgg cct cat agg ctc atg tag ctg cca atc att aat 1801
V C D L S F S Q R P H R L M tag * 448

gct ccc agc ctt cat ctc caa taa tat cca gaa tag ctt agg aca ttc cta ttg tgt gtg 1861
tgt gtg tgt tta tgt ctg aat gca ttg ttt caa tga ttg taa ctt att tta ttc aca caa 1921
acc tta ttt tta ctt gcc gat gta ggt ctt aat gca ctt ctt ttc tta ata tcc tta aag 1981
gag att cac tat tat act cta att gtt taa gaa gag gtt gtg caa tat att gta att 2041
taa gtg cta aag gat gat ctc agg aag atg cac att ctc 2080

```

Figure 4. Nucleotide and deduced amino acid sequences of goldfish *noc-aa*. Sequences are accessible through GenBank. Splice Variant I is composed of exon 1a in blue (coding sequence with cyan background), exon 2 (green background) and exon 3 (yellow background) (allele 1 accession no. OR651354). Splice Variant II is composed of non-coding exon 1b, exon 2, and exon 3 (accession no. OR651356). Red arrow indicates the insertion point of exon 1a or exon 1b on exon 2. Bold font indicates stop codons upstream of the initial methionine codon. Coding region extends from the first methionine residue (blue box for variant I, green box for variant II) to the stop codon (* and red box). Backgrounds colored cyan, green, and yellow indicate the coding part of exon 1a, 2, and 3, respectively. Red letters indicate differences in nucleotides and amino acids between the two allelic sequences (allele 2, accession no. OR651355 and OR651357).

```

(exon 1b)
ctg gat tag ttc ttt gct ccc tca aag a cta acc cac tga acg gac act gga tca gca 31
ctg gat tag ttc ttt gct ccc tca aag tca aag cga aga gga ttg tgc ctt taa atc aca 91
tga tca cat cgg aga caa gtc agt gga aaa gga taa aat tag agt gtg tca g↓ 143

(exon 1a)
tac gac taa acc att aat tgt tgt gta ggt gaa cat ogc gtt aaa aca tgt agc ttt ttt 60
ttt att aat gta cat atg att tga aat agt ttg aaa tgt tta aac cct aag tgt agt tgt 120
aat att tca tgt ata taa taa ctt aat aca tta tgc tat ttt att agt gca tga aaa tct 180
aat cta ctg aac tcc cct atc aag aat tgt cac gtt gga gta aga tga aac ttc tac cgg 240
cat ttt tcc ata caa taa aaa aaa aat caa oga cac gtc taa gca gag tca aac atg gct 300
ttt aca agt gtt aaa tag gtc tgt aag ctt att tct tgc tta gcc aaa ctt aac gcc att 360
cgt ctt tta tca gtt taa act gca gtt cct gtc gac cga aga aga atg taa cgg gct 420
M Y F A 4

ggg cgc tgt tca tog ctg ttt cac oga gac ctg gcc gcg gtc tgt etc tca tog ctg ggg 480
R R C S S L F H R D L A A V C L S S L G 24
aca cat cca aag aaa ccc ggt cgg ctg aag aaa agc agt ctt cgg agt tca cga gtt cat 540
T H P K K P G P L K K S S L P S S R V H 44
ttt cct cca aca agy tca aac tca tcc cca gtt ctg ctc agt tct tta caa gtg tgt caa 600
L P P T R S N S S P V L L S S L Q V C Q 64
atg ggc agc agc agc agc aga ctc ttc agt act gtg gcc caa act ctg agc agc gct gcc 660
M G S S S S R L F S T V A Q T L S S A A 84
ctg aga gac ctg gat gta gat act gat gat gat gag tat gag cag gct gac ggt gat ggc 720
L T D L H V D T D D D E Y E Q A D R D A 104
ctg ctg cgt gaa tgc gag gat gtt ttg agg aac cgg cct cct oga ctg cac oga gac ttc 780
L L R E C E D V L R N R P P R L H R D F 124
atg atg acc gga gcc tgc tca ctg cag agc gcc ccc atc ogc gtc atg cag tgg aac atc 840
M M T G A C S L Q S A P I R V M Q W N I 144
tta gca caa gct ctg ggt gaa ggc aag gat ggt ttt gtg ogc tgc cca atg gag gcg ttg 900
L A Q A L G E G K D G F V R C P M E A L 164
aac tgg tca gag agg aag tat ctc att ctg gag gag atc cta acc tac aag cca gat gtg 960
N W S E R K Y L I L E E I L T Y K P D V 184
ttg tgt cta cag gag gtg gac cac tac ttt gac acc tto cag cct gtg ctc gcc agc ttg 1020
L C L Q E V D H Y F D T F Q P V L A S L 204
ggc tac cag agc agc ttc tgc ccc aag cca tgc tcc ccc tgc ctg gac gtt cac aat aat 1080
G Y Q S S F C P K P C S P C L D V H N N 224
aac ggc ccc gac ggc tgc gct ctg ttc ttc aac ogc ogt ogc ttc cag ctg ctc cac acc 1140
N G P D G C A L F F N R R R F Q L L H T 244
gct cat ctc aga ctc tct gcc atg atg ctc aaa acc aac cag gtt gct atc gtg gct acg 1200
A H L R L S A M M L K T N Q V A I V A T 264
ctt ogc tgc agg ttc aca ggt oga gtc ttc tgc gtg gca gta aca cac ttg aaa gog cgg 1260
L R C R F T G R V F C V A V T H L K A R 284
agc ggc tgg gag gcg ttt oga agc gca cag ggg gct cac ctc ctc cag cag ctc tgt gaa 1320
S G W E A F R S A Q G A H L L Q Q L C E 304
cta acc aca cag tca cac aca gag atg gat gat gaa aac cgg act aga gtt ggg gaa acc 1380
L T T Q S H T E M D D E N R T R V G E T 324
aaa gac ggt att cog tta gtg gta tgt ggg gat ttt aat gcg gag cca agt gag gaa gtg 1440
K D G I P L V V C G D F N A E P S E E V 344
tat agg ogc ttt atg act tct cct ctg ggg cta gat agt gct tac aag tgt cta agt gag 1500
Y R R F M T S P L G L D S A Y K C L S E 364
gac ggg aac aca gag cca ccc tac act agc tgg aag atc cgg ccc agc gga gag agc ogc 1560
D G N T E P P Y T S W K I R P S G E S R 384
tcc acc ttg gac tac att tgg tac tca gaa aga gca ttt cag gtg gat gct gtg cta aag 1620
S T L D Y I W Y S E R A F Q V D A V L K 404
ata cca agc gaa gag cag atc ggg cog gac cgg ctt cog tog tat cat tat cct tct gat 1680
I P S E E Q I G P D R L P S Y H Y P S D 424
gat ctg tca ctg gtg tgt gac ctt agt ttt agt cag aag cct cat agg ttc atg Tag ctg 1740
H L S L V C D L S F S Q K P H R F M * 442

tca atc att agt gct ctc caa tag gga tcc aga gca gct tag gac att agg att atg tgt 1800
gca caa gtc tga atg tga caa tgc att att tgc atc att gca aag aat gta aca cat ttt 1860
att cat cta agc ctt att ttt acc tgc tga tgt agt tct taa tgc acc tct ttt att 1920
att ttc ttc gat gga gat taa tta tta tta tta tta tgc tca ctg ttt gtt taa cag agg 1980
ttg tac aat ata ttg taa ttt att aag tgc tag agy atg atc tca gga aga tgc aca ttc 2040
tog att aaa agg ttc cac aag tgg aaa gat gtt tat att tat act gac tgt aca ttt tct 2100
ttt ctt tta gta aca ctg taa tca ttg ata tgt ttg oca ttg acc agt aat ctc taa ata 2160
cat cag tgt att taa att gtg tgg tct ctc tct tta ata aca aac taa oga at 2213

```

Figure 5. Nucleotide and deduced amino acid sequences of goldfish *noc-ab*. Both sequences are accessible through GenBank. Splice Variant I is composed of exon 1a in blue (coding sequence with cyan background), exon 2 (green background), and exon 3 (yellow background) (accession no. OR651297). Splice Variant II is composed of non-coding exon 1b, exon 2, and exon 3 (accession no. OR651298). Red arrow indicates the insertion point of exon 1a or exon 1b on exon 2. Bold font indicates stop codons upstream of initial methionine codon. Coding region extends from the first methionine residue (blue box for variant I, green box for variant II) to the stop codon (* and red box). Backgrounds cyan, green and yellow indicate the coding part of exons 1a, 2, and 3, respectively.

We determined that both *noc-aa* and *noc-ab* paralogues have a very similar structure with many common features (Figures 4 and 5). The stop codon TAG terminates the coding sequence, insertions of introns are in phase 1, and the separation between exons is identical, being a GTG codon in the transition of exon 1–exon 2 and a GCT codon in the case of exon 2–exon 3. The initial methionines of the splicing variants I and II have the same relative position in the coding exon 1a for variant I (in cyan) and exon 2 for variant II (in green). We correctly verified the position of initial methionines, because of the presence of many in-frame stop codons upstream (bold letters in 5'UTRs, Figures 4 and 5).

2.2.2. Transcription of *Nocturnin bb (noc-bb)* Gene

We have identified two splicing variants (II and III) for the *noc-bb* paralogue (Figure 6), with the two corresponding starting ATG codons. Exons 2 and 3 are common to both variants (II and III), but the presence of a short coding exon 1c (13 pb, pink box in Figure 6) is specific to variant III. The corresponding sequence (MEVVA-) fulfils the requirement of the DEG Nend UBRbox 2 motif (Supplementary Figure S1).

The separation between exons is different compared to both *noc-a* paralogues, with a GCA codon in the two links between exons. In addition, the sequence of the stop codon is TAA, showing another difference from *noc-a* paralogues. For the *noc-bb* transcript, we reached the poly-A tail (not included in Figure 6).

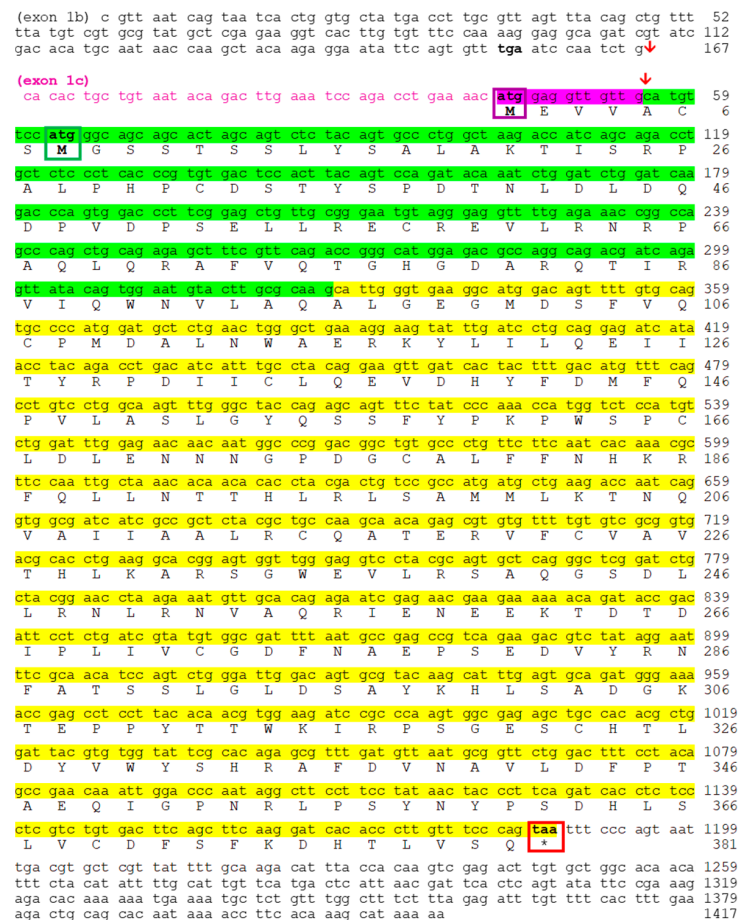


Figure 6. Nucleotide and deduced amino acid sequences of goldfish *noc-bb*. Both sequences are accessible through GenBank. Splice Variant III is composed of exon 1c in purple (coding sequence with purple background), exon 2 (green background), and exon 3 (yellow background) (accession no. OR651299). Splice Variant II is composed of non-coding exon 1b, exon 2, and exon 3 (accession no. OR651300). Red arrow indicates the insertion point of exon 1b or exon 1c on exon 2. Bold font indicates stop codons upstream of initial methionine codon. Coding region extends from the first methionine residue (purple box for variant III, green box for variant II) to the stop codon (* and red box). Backgrounds colored purple, green and yellow indicate exons 1c, 2, and 3, respectively.

2.2.3. Protein Alignment and Deduced Secondary Structure

Variants I of *noc-aa* and *noc-ab* are characterized by a well-conserved long first coding exon (60–65 amino acids), described as a mitochondrial-targeting signal (MTS) in mammals. We performed the analysis of variant I of cyprinid nocturnins, looking for MTS motifs by using MitoFates and Tppred 3.0 tools. Tppred is more sensitive than MitoFates for detecting

MTS, and our results confirm that the variant I contains a real MTS (Supplementary Figure S2), while variants II and III lack any trace of the MTS motif.

Alignment of the amino acid sequences deduced from the three paralogues of goldfish nocturnins with the matching fragments of the sequences from other vertebrates was performed (Figure 7). The splicing variant II was used for the three isoforms because it is the core sequence common to all vertebrates. The result reveals a high conservation of these sequences, especially in the central zone and the C-terminal region. In the alignment, we identified amino acids linked to the catalytic activity of the human NOC, verifying its conservation throughout the evolution, represented by green boxes in the alignment (Figure 7: Asn¹⁴⁹, Glu¹⁹⁵, Lys²¹⁹, His²⁸⁶, Lys²⁸⁸, Arg²⁹⁰, Asp³²⁴, Asn³²⁶, Arg³⁶⁷, Asp³⁷⁷ and His⁴¹⁴). The leucine zipper motif is a segment of 30 amino acids organized in an α -helix, with periodic repetition of leucines in every seventh position and their side chains line up on the same side of the α -helix. In the alignment, a region which resembles this structure can be observed, with some variations. The fourth leucine residue has mutated to threonine in cyprinids, the third leucine in goldfish NOC-AA is replaced by a methionine (a hydrophobic amino acid like leucine), and in the case of humans, the third leucine is replaced by a phenylalanine (also hydrophobic). Only *Xenopus tropicalis* retains this motif unchanged.

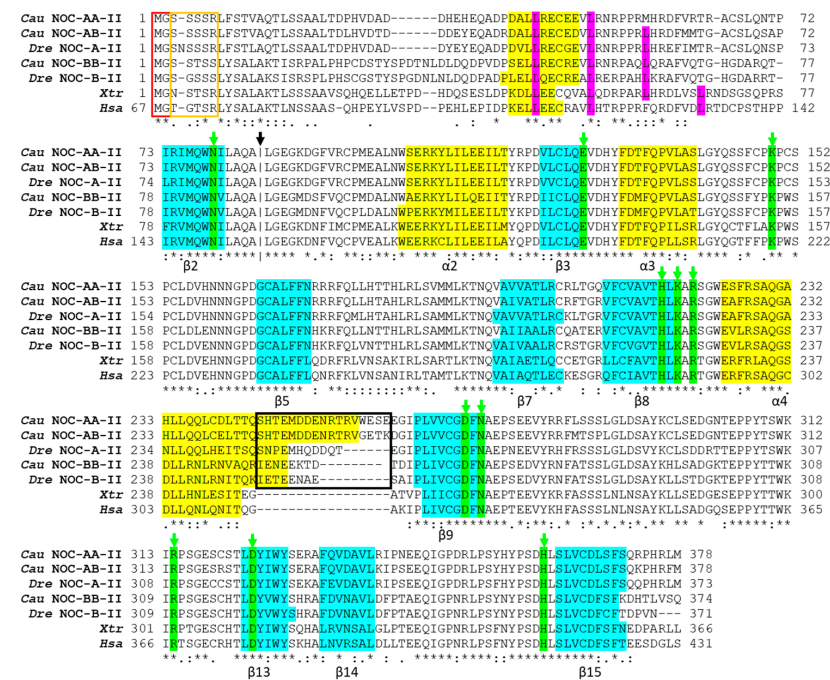


Figure 7. Alignment of the deduced amino acid sequences of splicing variant II of goldfish NOC-AA (Cau NOC-AA-II: WNX29031), NOC-AB (Cau NOC-AB-II: WNX29026), and NOC-BB (Cau NOC-BB-II: WNX29028) with NOC sequences from *Danio rerio* (Dre NOC-A-II: XP_005169676.1) and (Dre NOC-B-II: XP_021331689.1), *Xenopus tropicalis* (Xtr: KAE8630336.1), and *Homo sapiens* (Hsa: NP_036250.2). Multiple sequence alignment was conducted using Clustal v.X2. Bottom symbols (.:*), indicating a significant conservation from minor to major. Horizontal hyphens indicate gaps introduced to optimize the alignment. The vertical hyphen line (black arrow) marks the transition between exon 2 and 3. The secondary structure was predicted using JPred4. Structural elements are indicated as follows: yellow boxes (α -helix) and blue boxes (β -sheet). The names of α and β follow the nomenclature of Abshire and coworkers [27]. Green boxes pointed to by a green arrow indicate the amino acids conserved in the active site. Pink boxes mark the position of the heptad leucine repeat. Red rectangles indicate putative myristoylation sites, and the orange rectangle indicates a Ser/Thr-rich region. Black rectangles indicate the non-conserved domain of goldfish NOC paralogues compared to human and frog NOC.

Other conserved features of NOC proteins may be related to their function or sub-cellular location, as indicated by the presence of an N-myristoylation motif (MGXXXS) in the N-terminal position (red rectangle in Figure 7) and the nearby Ser/Thr-rich region (orange rectangle in Figure 7). All cypriniform NOCs (variant II) have a high confidence score from the Myristoylator program, higher than 0.91, except *Carassius gibelio* NOC-AB, *Cyprinus carpio* NOC-AB, and NOC-BA, in which the N-myristoylation motif was lost (Supplementary Figure S1).

A prediction of the secondary structure of the three NOC isoforms obtained with JPred4 is shown in Figure 7. Those results with values of low confidence and with less than three residues for either α -helix or a β -sheet are not indicated in the figure. We observed a total of four α -helices and nine β -sheets conserved in the three *noc* isoforms of goldfish.

2.2.4. Tertiary 3D Structure

The prediction of the tertiary structure of NOC proteins was obtained with the SWISS-MODEL program, which creates a possible tertiary structure from an aminoacidic sequence, taking an evolutionary-related protein as template. The crystallized structure of the human NOC catalytic domain has been used as a model (PDB accession number: 6bt1) for the prediction of the tertiary structure of the three splicing variant II of goldfish nocturnins. The human template was truncated (without the 119 N-terminal amino acids containing the MTS motif) to obtain the correct crystallization of the core domain. The results show a well-conserved 3D distribution of α -helix and β -sheet motifs (blue domains, Figure 8) in goldfish. Meanwhile, only the small peripheral loop (orange) does not preserve its spatial structure. These non-conserved regions, between α 4 and β 9, match with the sequences inside the black box of the protein alignment (Figure 7).

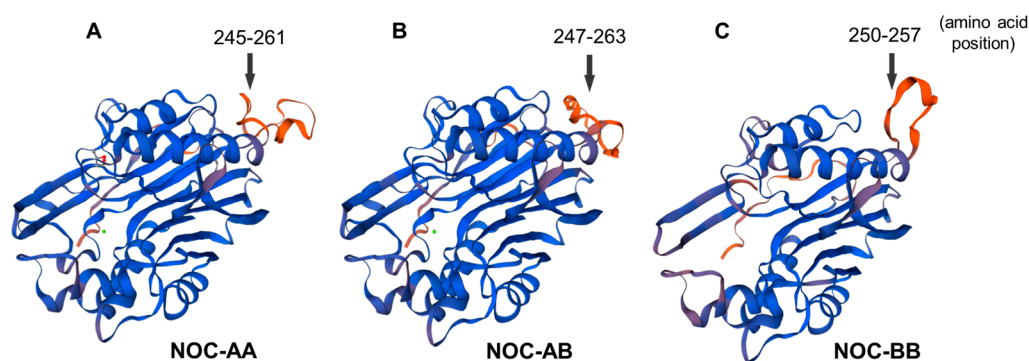


Figure 8. Tertiary predicted 3D structures of splicing variant II of goldfish; nocturnin: NOC-AA (A), NOC-AB (B), and NOC-BB (C). In blue are the conserved segments. The non-conserved domains in human and goldfish are represented in orange (the arrows indicate the amino acids' position in Figure 7).

2.3. Proposed Classification for Cyprinine Nocturnins

Our next objective was to establish a classification and a long-standing nomenclature of nocturnins throughout the teleost phylogeny. We took the zebrafish as a base, where a clear distinction between *noc-a* and *noc-b* exists, corresponding to specific teleost WGD or 3R. We performed a complete synteny analysis and created a phylogenetic tree of protein sequences.

2.3.1. Synteny Analysis

Synteny analysis allows us to determine the conservation of the closest loci of *noc* in goldfish and four Cyprinidae species: *Danio rerio* (zebrafish), *Sinocyclocheilus grahami* (golden-line barbel), *Cyprinus carpio* (common carp), and *Carassius gibelio* (Prussian carp). For this task, we used the publicly accessible database indicated in Supplementary Table S2. In addition, we used the comparative analysis of these five species to identify loci that can

be used to discriminate between matrilineal and patrilineal chromosomes, and that could be used as predictors in other Cyprininae species.

Figure 9 shows the synteny analysis of *noc-a* paralogues. There is a strict conservation of genes upstream and downstream of *noc-a*, from *ccdc88b* to *naa15a*, in all orthologous chromosomes. The paralogue *elf2a*, downstream neighbor of *noc-a*, may be used as a good predictor of *noc-a* paralogues' location in other cyprinid species. The *flrt1b* gene is inserted into an intron of the *macrod1* gene in the opposite direction of transcription. In Figure 9, *flrt1b* appears as a gap at the center of the *macrod1* gene.

The search for discriminant (matrilineal–patrilineal) loci allowed us to identify the *fgfbp1b* gene downstream of *noc-aa*, and an uncharacterized gene between *scyl1* and *ccg8* upstream of *noc-aa*, which are specific and diagnostic patrilineal markers. A duplication of *rarres3* in *Carassius gibelio* is conserved in *C. auratus*. The respective matrilineal chromosome of cyprinines is characterized by the loss of several genes, such as the *fgfbp1b* gene located between *anxa5a* and *fgfbp2a* in zebrafish. Furthermore, there is a point deletion of *ugl*, placed between *naa15a* and *ra33ba*, and the *btbd18* deletion located between *cryba111* and *tmem33*; both deletions are shared by *C. gibelio* and *C. auratus*.

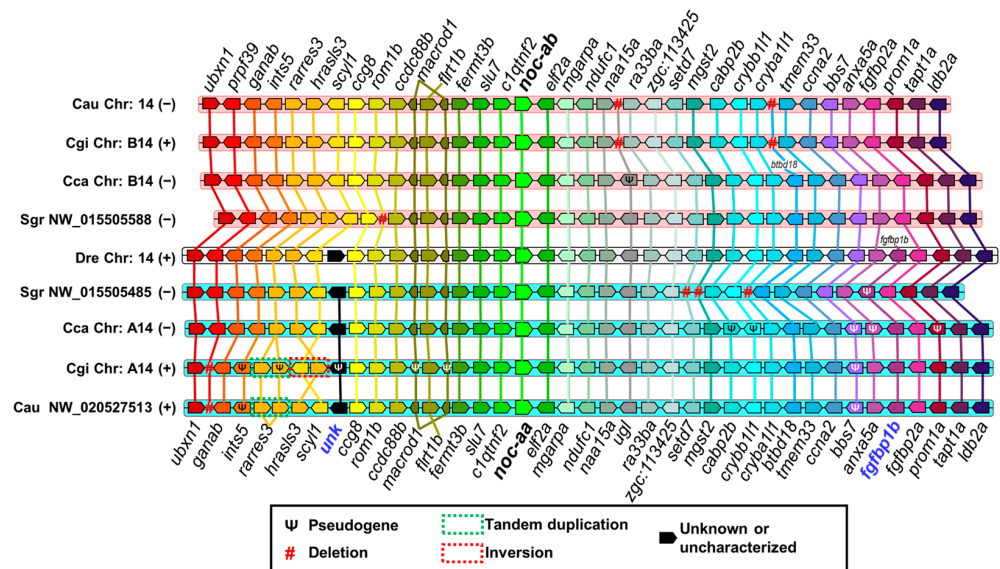


Figure 9. Synteny analysis of *noc-a* paralogues in *Danio rerio* (Dre), *Sinocyclocheilus graham* (Sgr), *Cyprinus carpio* (Cca), *Carassius gibelio* (Cgi), and *Carassius auratus* (Cau). Black bolded names indicate nocturnin paralogues in the center of synteny. Chromosomes are represented with a white background in *Danio rerio* (as the reference species) and a cyan or pink background in the other species, depending on the patrilineal or matrilineal subgenome, respectively, of the ancestral duplication of Cyprininae (4Rc). Chromosome coding corresponds to that of the genome project (Supplementary Table S2), and the (+) or (–) symbols represent the sense or antisense orientation of the chromosome, respectively, in the projects analyzed. Genes are represented with pentagons whose sharp vertex indicates the transcription sense. Same colors indicate orthologous genes. Black pentagons represent unknown genes or genes without orthologous in any chromosome analyzed. Mutations are inferred comparing the orthologous chromosomes: pseudogene (Ψ), deletion (#), tandem duplication (green dashed boxes), and inversion (red dashed boxes). Gene abbreviations are indicated in the Supplementary Table S3. Gene names in bold blue indicate diagnostic markers of patrilineal chromosomes.

Figure 10 shows the synteny analysis of *noc-b*. The gene *elf2b* is located downstream of *noc-b* in all the orthologous chromosomes. Upstream of *noc-b*, chromosomes show a large region with no genes annotated up to *pcdh10a*. Chromosomes in Figure 10 are not to scale, and this gap is not represented. The chromosome of the golden-line barbel is represented in two separate segments, because the assembly of its genome is incomplete.

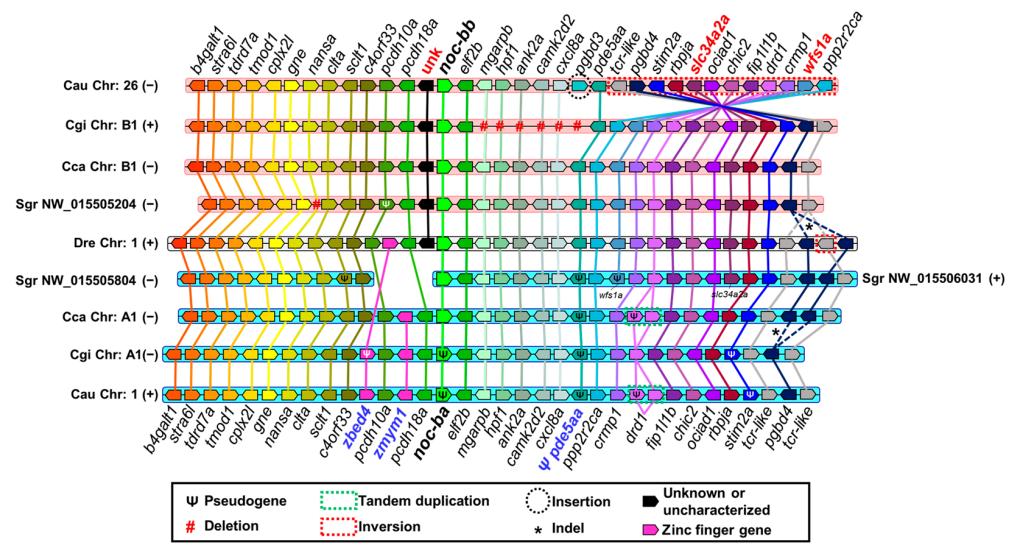


Figure 10. Synteny analysis of *noc-b* paralogues in *Danio rerio* (Dre), *Sinocyclocheilus graham* (Sgr), *Cyprinus carpio* (Cca), *Carassius gibelio* (Cgi), and *Carassius auratus* (Cau). Black bolded names indicate nocturnin paralogues in the center of synteny. Chromosomes are represented with a white background in *Danio rerio* (as reference species) and with a cyan or pink background in the other species depending on the patrilineal or matrilineal subgenome, respectively, of the ancestral duplication of Cyprininae (4Rc). Chromosome coding corresponds to that of the genome project (supplementary Table S2), and the (+) or (−) symbols represent the sense or antisense orientation of the chromosome, respectively, in the projects analyzed. Genes are represented with pentagons whose sharp vertex indicates the transcription sense. Same colors indicate orthologous genes. Black pentagons represent unknown genes. Pink pentagons indicate zinc finger genes. Mutations are inferred comparing the orthologous chromosomes: pseudogene (Ψ), deletion (#), tandem duplication (green boxes), inversion (red boxes), insertion (black circle), and provisional indel (*). Gene abbreviations are indicated in the Supplementary Table S4. Gene names in bold blue (below) or bold brown (above) indicate diagnostic markers of patrilineal or matrilineal chromosomes, respectively.

The map shows the presence of *wfs1a* and *slc34a2a* downstream from *noc-bb*, and an uncharacterized gene upstream from *noc-bb* as adequate markers of matrilineal origin of these chromosomes. In *Carassius gibelio*, it is important to consider a deletion of a cluster between *elf2b* and *pde5aa*. However, in goldfish, the insertion of *pgbd3* gene and an inversion of loci from *tcr-like* to *ppp2r2ca* is more representative. Regarding the patrilineal chromosomes, wherein *noc-ba* resides, we have determined the presence of two zinc finger genes upstream of *noc-ba*, with *zmym1* being conserved in common carp, Prussian carp, and goldfish, and *zbed4* in species of *Carassius*. The absence of this segment in the golden-line barbel genome impedes their accurate identification. In all the cyprinines analyzed, pseudogenization may be a frequent process in these patrilineal chromosomes, and we can use it as a diagnostic tool. For example, *pde5aa* in all species and *stim2a* and *noc-ba* in *Carassius* species are pseudogenized with the accumulation of several deleterious mutations. In addition, the *drd1* gene has tandem-duplicated in goldfish and common carp with a pseudogenized copy. Prussian carp (*C. gibelio*) has lost completely this pseudogene.

To quantify the similarity of two synteny, we proposed two quantitative parameters: the synteny index (SI, related to the number of relative positions conserved) and the synteny conservation rate (SCR, as a percentage of maximum theoretical synteny conservation). These numerically indicate the conservation of relative gene positions in two species. The detailed definitions and calculation methods of both parameters are in Appendix A.

We compare goldfish with Prussian carp, common carp, and zebrafish, according to its decreasing phylogenetic distance (Figure 11C). We exclude golden-line barbel due to the fragmentary status of its genome project. In the orthologous comparison (matrilineal to matrilineal, or patrilineal to patrilineal), the SI is higher than 1 (Figure 11A) and the SCR is

over the 50% (Figure 11B) for all *noc* paralogues, indicating that the closest loci to *noc* are very similar in Cyprinidae. As expected, when goldfish and zebrafish were compared, both parameters show low values, in agreement with their larger phylogenetic distance (different subfamilies). The comparison zebrafish *noc-b* vs. goldfish *noc-ba* has an exceptionally low synteny conservation (SI: 1.4 and SCR: 59%), because of many mutations (inversions and indels of individual genes) on goldfish chromosome 1. They have a more marked effect in reducing synteny indexes than the single large inversion on chromosome 26.

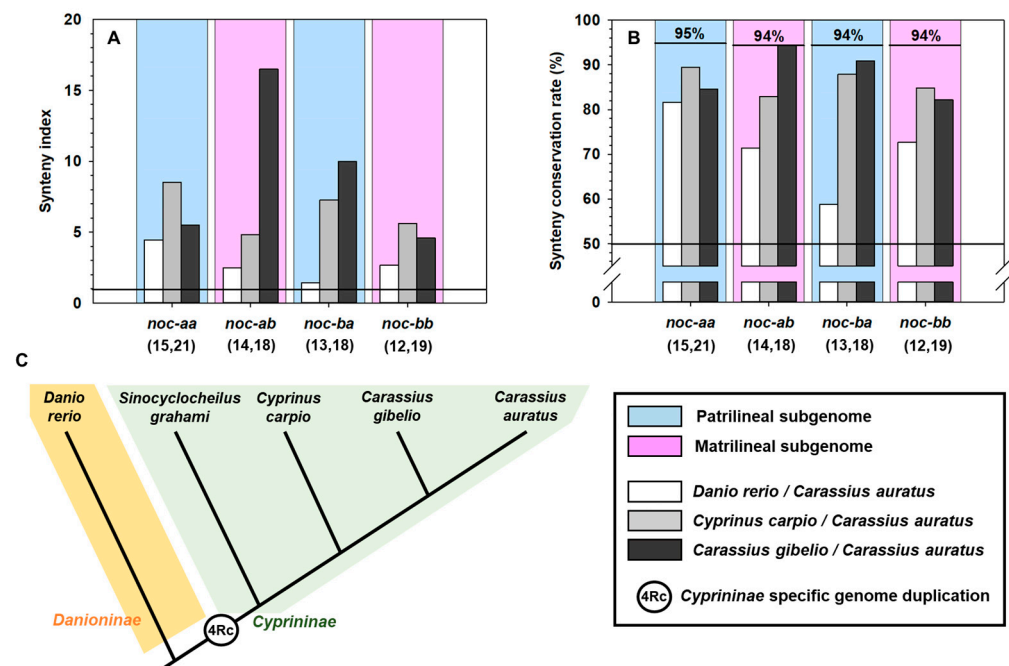


Figure 11. (A,B) Synteny index and synteny conservation rate in *Carassius auratus* compared to *Danio rerio* (white), *Cyprinus carpio* (light gray) and *Carassius gibelio* (dark gray). Numbers in parentheses indicate genes upstream and downstream of the *noc* paralogue analyzed. Blue and pink backgrounds indicate patrilineal or matrilineal origin of chromosomes, respectively. The lower horizontal line indicates the threshold (1 for synteny index, 50% for synteny conservation rate), where the conserved loci positions are more frequent than the non-conserved. The line above in the graphs in B indicates the maximum synteny conservation rate (%) for each paralogue. (C) Assumed phylogeny of species analyzed in the synteny study.

In Cyprininae, *noc-ab* and *noc-ba* synteny conservation follows the expected genetic distance, i.e., higher values for Prussian carp vs. goldfish than the common carp vs. goldfish comparison. Furthermore, *noc-ab* reached the maximum synteny conservation rate (mSCR, Appendix A) of 94% in both *Carassius* species, indicating the perfect synteny of these chromosomal segments (Figure 9). In *noc-aa* and *noc-bb*, we found that the synteny conservation is higher for common carp vs. goldfish than for Prussian carp vs. goldfish. This paradoxical result may be explained by the evident changes in Prussian carp chromosomes, as the multiple loci deletion in chromosome B1, and the inversion of *scyl1-hrasls3* segment in chromosome A14.

The high values of these two numerical parameters are only obtained from the orthologous contrast test, meaning comparison of matrilineal–matrilineal or patrilineal–patrilineal chromosomes (Figure 11). In the ohnologous contrast test (comparison of matrilineal to patrilineal and vice versa), the synteny index drops by two to ten times, and synteny conservation rate is reduced from a half to a third (Supplementary Table S5).

2.3.2. Phylogenetic Tree

A phylogenetic tree was designed with the amino acid sequences of the NOC splicing variant II of each paralogue, except for *H. sapiens*, *Sinocyclocheilus rhinocerosus*-BA, *S. anshuiensis*-AA and AB, *S. grahami*-AA, and *Cyprinus carpio*-AB, where the N-truncated form of NOC was used to eliminate exon 1a of splice variant I, when variant II was not available in GenBank. Figure 12 shows the result obtained by the ML method. The NOC from Sarcopterygii (*H. sapiens*, *X. tropicalis*, *Latimeria chalumnae*) and the ancestral Osteichthyes (*Lepisosteus oculatus*) appears as a single gene in the basal position of the tree. Teleosts nocturnins are divided into two main clades (NOC-A and NOC-B) with a mean of 60% identity. This gene duplication corresponds to teleostean 3R. As expected, in the most complete teleostean genomes available, NOC-A and NOC-B always resides in different chromosomes. *Lepisosteus oculatus* has a single *noc* gene, which has not undergone the 3R duplication. To assess the nomenclature A or B, we follow the proposed for zebrafish nocturnins.

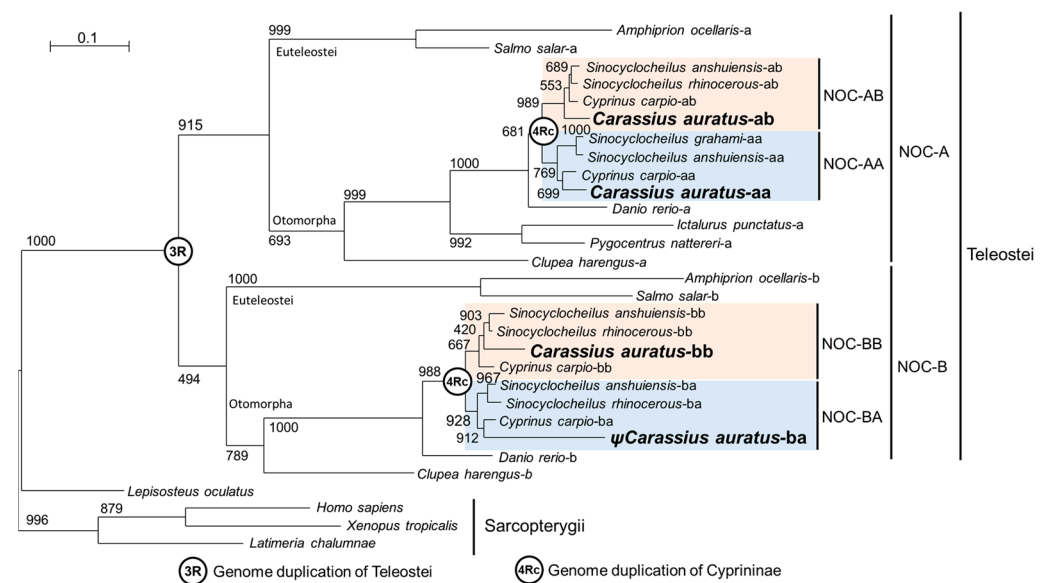


Figure 12. Phylogenetic tree showing the 4Rc relationships among NOC sequences. The evolutionary model used was the Jones–Taylor–Thornton, Gamma-distributed (JTT + G). The tree was inferred by the maximum likelihood method (ML). The numbers in the nodes refer to bootstrap values of a total of 1000 replicates. The scale bar indicates the average number of substitutions per position. The binomial name of the species is given on the right side of the tree. Letters A and B indicate NOC isoforms in teleosts. 3R and 4Rc indicate the proposed whole-genome duplication events in Teleostei and Cyprininae, respectively. Blue and pink boxes indicate the patrilineal or matrilineal clades of Cyprininae species, respectively. Ψ , indicates pseudogene. Species names and GenBank accession numbers of the sequences are indicated in Supplementary Figure S3.

The tree reveals the main division of Clupeocephala (Otomorpha and Euteleostei) as monophyletic branches for both NOC-A and NOC-B. On the other hand, the phylogenetic analysis makes it possible to locate the full genome duplication of the Cyprininae subfamily (4Rc) in the tree, with moderate bootstrap support (above 667). From this node, we can observe a symmetrical bifurcation point with representatives in each branch of all cyprinine species analysed. We choose the branch nomenclature AA, AB, BA, and BB, to show the allotetraploid origin of cyprinines, and the patrilineal (genome A) or matrilineal (genome B) subgenomes (Figures 9 and 10). The close relationship of the NOCs inside each branch (identity of about 90% for AA-AB branch and 88% for BA-BB branch), in all cyprinines (*Sinocyclocheilus* sp., *Cyprinus* sp., *Carassius* sp.) allows us to verify that the names assigned to the *noc* are correct. The goldfish NOC-BA paralogue is connected to the NOC-BA cluster with a long branch caused by the high number of mutations in this pseudogene. However, NOC-BA from *Cyprinus* and *Sinocyclocheilus* are functional genes

and are located in the expected position in the tree, with a similar branch length compared to other nocturnin paralogues.

2.4. Tissue Expression of *noc* mRNA in Goldfish and Zebrafish

The relative gene expression of *noc-aa*, *noc-ab* and *noc-bb* in goldfish is shown in Figure 13. The three paralogues have a ubiquitous expression, with a distinctive pattern for each paralogue. In peripheral tissues, the adipose tissue showed the highest expression of the three *noc* paralogues, with significant lower values in the spleen, hepatopancreas, skin and muscle (Figure 13A,C,E). The lowest expression of these paralogues was found in the gastrointestinal tract.

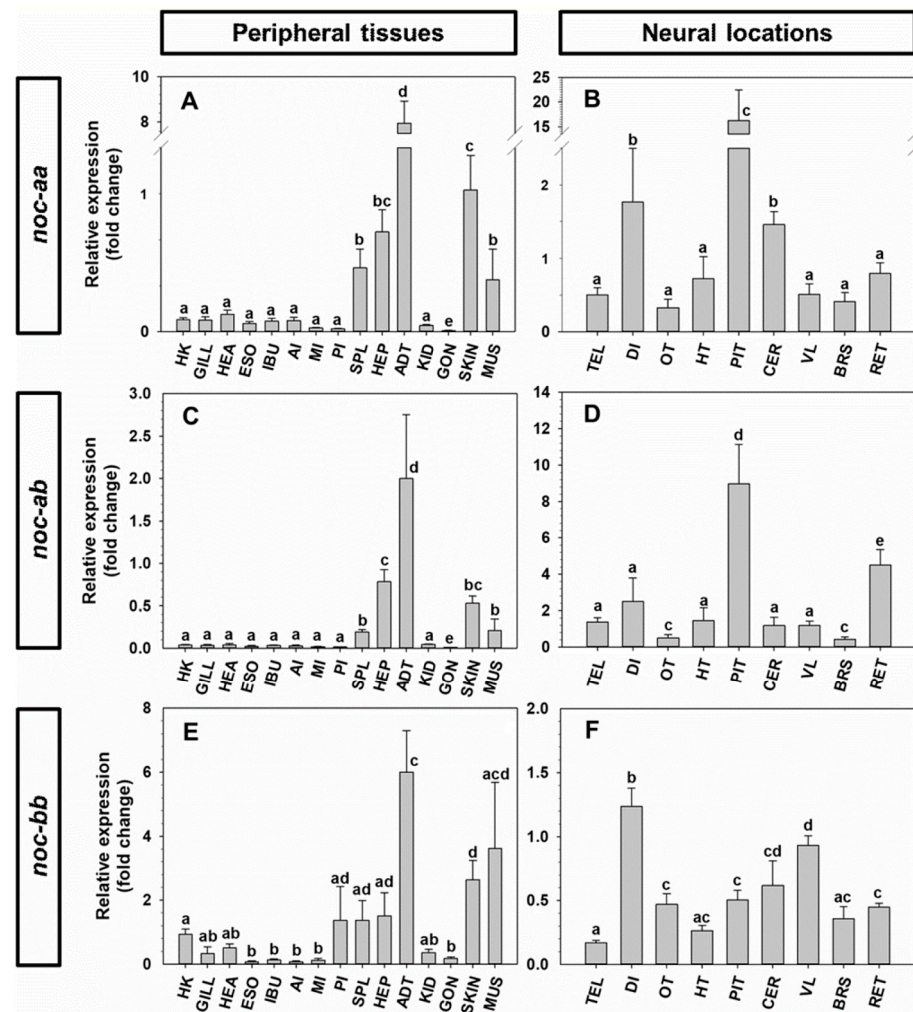


Figure 13. Expression of *noc-aa*, *noc-ab* and *noc-bb* in peripheral (A,C,E) and neural tissues of goldfish. Data are expressed as mean + S.E.M. ($n = 6$) relative to the hepatopancreas (A,C,E) and to the diencephalon (B,D,F). Different letters indicate statistical differences among tissues (ANOVA-SNK; $p < 0.05$). Head kidney (HK), gill (GILL), heart (HEA), esophagus (ESO), intestinal bulb (IBU), anterior intestine (AI), middle intestine (MI), posterior intestine (PI), spleen (SPL), hepatopancreas (HEP), adipose tissue (ADT), caudal kidney (KID), gonad (GON), skin (SKIN), muscle (MUS), telencephalon (TEL), diencephalon (DI), optic tectum (OT), hypothalamus (HT), pituitary (PIT), cerebellum (CER), vagal lobe (VL), brainstem (BRS), retina (RET).

In nervous system, both *noc-aa* (Figure 13B), and *noc-ab* (Figure 13D) are highly expressed in the pituitary in contrast to the other brain areas. Moreover, *noc-aa* (Figure 13B) transcripts are twice as abundant in the diencephalon and cerebellum with respect to the telencephalon, optic tectum, hypothalamus, vagal lobe, brainstem, or retina. However,

noc-ab transcripts are four times more abundant in the retina than in the telencephalon, diencephalon, hypothalamus, cerebellum, and vagal lobe; meanwhile, the brainstem and optic tectum showed the lowest abundance (Figure 13D). The expression of *noc-bb* is quite similar within the brain, with the highest abundance of transcripts in the diencephalon.

The relative abundance of the three *noc* paralogues in each of the tissues is represented in the Supplementary Figure S4. The gills, heart, esophagus, intestinal bulb, middle and posterior intestine, spleen, caudal kidney, skin, telencephalon, and hypothalamus showed similar amounts of the three paralogues. However, *noc-bb* is the most abundant paralogue in the head kidney, muscle, and gonads (up to 14 times, Figure S3A) and in the optic tectum, vagal lobe, and brainstem (Figure S3B). By contrast, *noc-aa* and *noc-ab* transcripts are 2–3 times more abundant than *noc-bb* in the anterior intestine, while *noc-ab* is 2–3 times more highly expressed than *noc-aa* or *noc-bb* in the hepatopancreas and the retina.

The relative abundance of *noc-a* and *noc-b* in the peripheral tissues, brain, and retina of zebrafish is summarized in Figure 14. The *noc-a* transcripts are highly expressed in the liver and pancreas (25–50-fold) compared to the brain and most of peripheral tissues, except for muscle (4-fold) (Figure 14A). The abundance of *noc-b* transcripts is more homogenous among tissues, with the highest amounts in the liver, gonads, muscle, and retina (Figure 14B); intermediate amounts in the pancreas, adipose tissue, and gonads; and significantly lower amount in the rest of the tissues. Regarding the relative expression of *noc-a* vs. *noc-b* in each of the tissues, similar amounts of both paralogues were found in the head kidney, gills, esophagus, posterior intestine, and the spleen (Supplementary Figure S4C). However, the intestinal bulb, anterior intestine, adipose tissue, caudal kidney, gonad, skin, muscle, brain, and retina were enriched in *noc-b* expression, while the *noc-a* was the most expressed paralogue in the liver and pancreas (Supplementary Figure S4C).

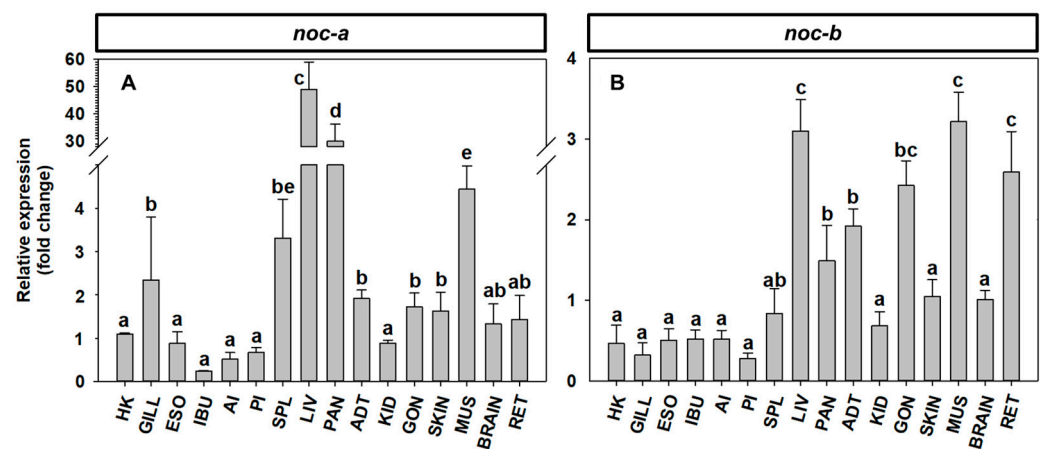


Figure 14. Expression of *noc-a* (A) and *noc-b* (B) in peripheral tissues and brain in zebrafish. Data are expressed as mean + S.E.M. ($n = 5$) relative to the brain. Different letters indicate statistical differences among tissues (ANOVA-SNK; $p < 0.05$). Head kidney (HK), gill (GILL), esophagus (ESO), intestinal bulb (IBU), anterior intestine (AI), posterior intestine (PI), spleen (SPL), liver (LIV), pancreas (PAN), adipose tissue (ADT), caudal kidney (KID), gonad (GON), skin (SKIN), muscle (MUS), brain (BRAIN), retina (RET).

3. Discussion

This work presents the first complete description of the splicing variants of the different *nocturnin* paralogues in cyprinid fish. To date, there are only two published studies in fish about this enzyme from partial gene sequences. In zebrafish, two isoforms of *noc* (*noc-a* and *noc-b*) in the retina have been characterized [5], establishing the basis for the nomenclature of this gene in the rest of the teleosts. In goldfish, *noc* expression has been related to feeding and fasting in the hypothalamus, hepatopancreas, and intestinal bulb [6].

Our present *in silico* analysis of *noc* genes from a primitive Osteichthyes, the spotted gar (*Lepisosteus oculatus*), indicates the possibility of an alternative splicing mechanism with three variants (Figure 1: I, II, III), which leads to a high complexity of gene expression, with functional specialization of each of the isoforms. If this hypothetical specialization confers biological advantages, it should be present in other vertebrate groups. Currently, two variants (I and II) can be identified in many genomes of Chondrichthyes, Osteichthyes, and Tetrapoda. However, we can only localize the three variants I, II and III in a few basal Osteichthyes, as coelacanth (*Latimeria chalumnae*: XP_005996687.1, XP_005996691.1, XP_005996689.1) and American paddlefish (*Polyodon spathula*: XP_041133185.1, XP_041133209.1, XP_041133201.1) in addition to spotted gar. Intriguingly, human nocturnin is an exception, with only a splice variant I.

It has been demonstrated that a teleost common ancestor underwent a full genome duplication (3R) 350 million years ago [24,28], generating the duplication of many genes, including *noc* with two paralogues: *noc-a* and *noc-b*. Due to this origin, both *noc-a* and *noc-b* are always found in separate chromosomes. After the additional genomic duplication of the Cyprininae subfamily (4Rc), estimated around 10–15 million years ago [26,29], four *nocturnin* genes were expected in separate chromosomes. We examined genomes from several allotetraploid species of *Sinocyclocheilus*, *Carassius*, and *Cyprinus carpio*, and we found two orthologous genes for zebrafish *noc-a* and two for *noc-b*. Thus, it is important to create a consistent nomenclature for these four paralogues in the Cyprininae subfamily.

The present results report the complete coding sequences (cds) of three *nocturnin* paralogues, allowing us to differentiate between *noc-aa* and *noc-ab* in goldfish. In other cyprinines, there are also two orthologous of the zebrafish *noc-b*, but only the *noc-bb* paralogue has been identified in goldfish by RT-PCR. After the duplication process, several mechanisms may preserve both copies, such as subfunctionalization and neofunctionalization. However, when a duplicated gene acquires a very specialized function, the most likely scenario is that one of the copies would be lost [28]. This is the case for *noc-ba* in the genus *Carassius*, where before total deletion, this gene accumulates deleterious mutations that hinder its transcription and subsequent translation into protein. Searching the complete genome of *Carassius auratus* we have found a sequence with high homology with the *noc-ba* of other cyprinines. However, the sequence shows several deletions in the coding regions (exons 2 and 3) and some insertions (exon 3). Therefore, we suggest that *noc-ba* could be a pseudogene that has accumulated deleterious mutations that changed the reading frame, and subsequently stopped transcription into functional mRNA. In addition, the genome of Prussian carp (*C. gibelio*) shows a highly mutated *noc-ba* (Figure 3), indicating that the process of pseudogenization started in the ancestor of both species.

In addition, to obtain the cds of the three isoforms of nocturnin, in this work, we aim to confirm the expression of the possible splicing variants of each paralogue, following the model of spotted gar *noc*. The common teleost ancestor, prior to 3R duplication, generated three possible transcripts for the first exon by alternative splicing, i.e., variants I, II and III. Some of these splicing variants have been conserved in the teleost evolution, and others have been lost. Our results, based on mRNA sequencing, demonstrate the splicing variant I and II for the two *noc-a* isoforms, (Figures 4 and 5), while II and III variants have been maintained for the *noc-bb* form (Figure 6). Nevertheless, failure to detect the variant III for *noc-aa* and *noc-ab* and the variant I for *noc-bb* may be due to a possible low expression in the sequenced tissue (hepatopancreas). To exclude this possibility, searching in goldfish transcriptomic SRA libraries obtained from several tissues (Supplementary Table S1), we found that the most abundant is the variant I for *noc-aa* and *noc-ab*; meanwhile, the forms II and III are a thousand times less abundant. The main variant for *noc-bb* is II, with the form III being minimal. Oddly, we found a very low, but significant, presence of the variant II of pseudogene *noc-ba* in SRA libraries.

Each splice variant needs a different codon ATG as translation initiation sites, but none of them have a canonical Kozak consensus sequence (5'-gccRccATGG-3', see Figures 4–6)

This event has been described for mammalian NOC, and several isoforms could be generated from the same mRNA by leaky translation initiation [30].

Comparing our sequences (comet goldfish variety) with those of the *Carassius auratus* (strain Wakin) genome in the NCBI database, we found some differences, mainly SNPs, which cause a few changes in protein sequences, as is expected when comparing sequences from different individuals of the same species. The NOC-AB proteins are identical, while the NOC-BB differs in one amino acid. By contrast, the sequence of NOC-AA seems to be more variable, as we found five different amino acids scattered in the protein sequence, and the specimen we used for sequencing was heterozygotic for *noc-aa*, with two alleles differing in four amino acids (Figure 4).

We note that *noc-aa* from the NCBI genome showed an insertion of two extra Ser at the beginning of variant II, with respect to our sequence WNX29031. This location is a hypervariable region for cyprinine NOC-AA and NOC-AB, with three to seven serial Ser repeats (Supplementary Figure S1) with an unknown functional meaning. This corresponds to variable tandem repeats of codons AGC or AGT. Insertion or deletion (indels) of repeat triplets is a highly probable event in some genes [31,32].

The well-preserved presence of multiple splicing forms for fish nocturnins makes us look for the possible functional role of each variant. We searched for specific sequence motifs at the N-terminal region. In variant I of mammalian *noc*, there is a putative mitochondrial-targeting signal (MTS) [30,33], which directs the nascent NOC chain into its final mitochondrial location. Once there, NOC suffers an intramitochondrial MTS elimination to release the core catalytic domain of enzyme [34]. The analysis carried out in this report allowed us to confirm that all variant I of cyprinids and most fish contain a real MTS of 76–90 amino acids (Supplementary Figure S2), with an initial amphipathic α -helix with net positive charge, followed by a disorganized region, and ending with an α -helix containing the appropriate cleavage sites for mitochondrial peptidases.

The variant II is the shortest NOC sequence, and it has a highly conserved N-terminal motif (MGXXS) in fish. This is the consensus sequence for the N-terminal glycine myristoylation. The co-translational mechanism involves the cleavage of N-terminal methionine from newly formed peptide by methionine aminopeptidases, and the addition of myristoyl group by cytoplasmic N-myristoyltransferases to now free glycine residue. This modification allows interaction with cell membranes and directs the final location of variant II. In mammalian cells, this cytoplasmic form of NOC is enriched on the outer surface of endoplasmic reticulum membranes [30].

The sequence of variant II is well conserved in most vertebrates, but we found a hypervariable region after an initial MG motif in cyprinid *noc-aa* and *noc-ab* (Supplementary Figure S1). Due to the proximity to myristoylation site, the number of serial serines in this hypervariable region may affect the myristoylation efficiency. Considering scores from Myristoylator tool, we can infer that the optimal serine number is 4 to 5, with scores higher than 0.97. Lower or higher Ser repeats significantly reduced the score. The sequence from goldfish genome (strain Wakin) shows a huge reduction in score (six serines: score -0.3432) with respect to our sequence of *noc-aa II* from the comet variety (WNX29031, four serines: score 0.9867). Then, in addition to the differences among species, there are also differences between specimens with functional consequences.

Variant III has a very short coding exon 1c of only 5–7 amino acids, which makes it difficult to detect by automated annotation of genomes, and only some fish species have it well annotated. Initially, we discard the variant III for *noc-a* paralogues in Cyprinidae, but a variant III of *noc-ab* is annotated in the Prussian carp genome (accession no. XP_052429539.1). A detailed search in the intronic sequences upstream from exon 2 allows us to detect this short exon, located at 2000–3000 bp upstream from exon 2 in all cyprinine species. The presence of this short coding exon overrides the N-myristoylation signal in variant III and creates an N-terminal motif M(D/E) detected by ELM software (Supplementary Figure S1). This is part of the DEG Nend UBRbox 2 motif, related to its turnover rate by targeting proteins for ubiquitin-dependent proteasomal degradation. The methionine aminopeptidases cleave

methionine and expose the negatively charged amino acid (Glu or Asp) to arginylation by arginyl transferases. Then, the positively charged N-Arg peptide is rapidly ubiquitinated and degraded [35].

This splicing variant III may assume a central role in the circadian function of nocturnins, as a putative clock-controlled gene [8]. To accomplish this role, it is required that NOC is a rapidly inducible, short-lived protein. In fact, classical clock-controlled genes, such as arylalkyl N-acetyltransferase, which is the rate-limiting enzyme of the daily rhythm of melatonin, have high turnover rates [36]. The first published NOC sequence (accession no. AAB39495), which was proposed as a clock-controlled gene in *Xenopus laevis* retina [1], is difficult to classify as homologous to variant I, II or III. It has an exon 1 of 21 amino acids with no MTS motifs, but it has a DEG Nend UBRbox 2 with an N-terminal (MD-) motif (Supplementary Figure S1). Thus, it may be the functional equivalent of a variant III in fish, justifying the fast and huge circadian rhythm of its expression in the frog retina, which shows an increase in mRNA from minimum to maximum value in only 4 hours [1], and a drop in protein from maximum to minimum in 4 hours [12].

For the core catalytic domain, most of it located in exon 3 and present in all splice variants, we verified the conservation of residues involved in Mg^{2+} -dependent phosphatase activity for goldfish, zebrafish, human, and frog NOC (Figure 7). Taking human NOC coordinates as reference, we found the conservation of the residues involved in the two Mg^{2+} bindings, for site 1 (Asn¹⁴⁹, Glu¹⁹⁵) and site 2 (Asp³²⁴, Asn³²⁶, Asp³⁷⁷) [27]. Two lysines (Lys²¹⁹, Lys²⁸⁸) involved in the binding of a pyrophosphate group from NADP(H) substrate, and two histidines (His²⁸⁶ and His⁴¹⁴ near Mg^{2+} site 2), as catalytic residues, are also conserved [15]. Homologous sites of all these conserved residues are also found in other members of the CCR4 family of the EEP superfamily (Exonuclease, Endonuclease, Phosphatase) [27]. However, we can point to residues only conserved in vertebrate NOC, goldfish included, but not in other CCR4 members, such as Arg²⁹⁰ and Arg³⁶⁷ (involved in the binding of the adenosine moiety of substrate NADP(H)), which allow the substrate selectivity of NOC [15].

The presence of a leucine-rich repeat motif (LRR) as a component of the multiprotein complex CCR4-NOT, belonging to the same superfamily as NOC [37], has been demonstrated. Its presence has been implicated in mediating protein–protein interactions [38]. In NOC, an LRR motif homologous to CCR4 does not exist [39]. However, we found four leucines in exon 2 with heptad symmetry forming part of a α -helix, characteristic of leucine zipper [40]. The complete conservation is only observed in *Xenopus tropicalis* NOC, although some variations appear in the other analysed species. Nevertheless, despite the variations in this leucine zipper in goldfish NOC sequences, the secondary structure prediction shows an α -helix in that region, indicating that it is possible for NOC to maintain the ability to interact with other proteins [7]. In this sense, mouse NOC binds and promotes the nuclear translocation of PPAR γ , but binding is through a nocturnin region far from the leucine zipper [20].

Allopolyploid organisms have an advantage over autopolyploids because chromosomes of patrilineal and matrilineal origin will be maintained as separate and stable evolutionary units (without meiotic recombination between them); therefore, they will maintain fertility despite the initial negative charge of polyploidy. It is necessary to consider that goldfish is an allotetraploid species ($4n = 100$), as are the rest of the cyprinines. The origin of this genomic duplication seems to be a hybridization of two ancestral species from different genera. The matrilineal subgenome (subgenome B) seems to come from a species near *Spiniothorax* [29] or *Puntius* [26], while the patrilineal subgenome (subgenome A) comes from a species of tribe Probarbini [26]. Regarding the chronology of these genomic events, and based on the molecular clock, it is assumed that the divergence between the *Danio* genus and the rest of the cyprinines took place 44 million years ago, the divergence between patrilineal and matrilineal ancestors of cyprinines 15 million years ago, and the process of allotetraploidization 13.7 million years ago [29].

Due to the recent origin of the allotetraploid cyprinids, it is likely that the sequences have not accumulated many differences, meaning it is not possible to determine a classification based only on sequence similarity. Therefore, we used two criteria to develop our classification: a phylogenetic tree based on coding sequences, and a synteny conservation analysis. These two criteria have shown a high phylogeny–synteny correlation for all *nocturnin* paralogues in the different species studied.

The chromosome naming of *Cyprinus carpio* and *Carassius gibelio* obeys the allotetraploid origin; they are divided in two sets of 25 chromosomes (A1 to A25, and B1 to B25 following the division of genome in the subgenomes A and B). The *Carassius auratus* genome too, although its chromosome nomenclature is different, has matrilineal chromosomes from 1 to 25, and a patrilineal subgenome from 26 to 50 (e.g., a pair of 4Rc ohnologous chromosomes are 1 and 26).

We expect that the four paralogues, *noc-aa*, *noc-ab*, *noc-ba* and *noc-bb*, will be located on goldfish chromosomes 39, 14, 26 and 1, respectively. However, we identified some deficiencies in the goldfish chromosome numeration: *noc-aa* is on an unplaced segment (NW_020527513, Figure 9) due to the incomplete assemblage of chromosome 39, but with a high homology with the corresponding patrilineal A14 chromosome from *C. carpio* and *C. gibelio*. In addition, following the expected orthology, goldfish *noc-bb* should be in the matrilineal chromosome 1, and *noc-ba* in the patrilineal chromosome 26, but their actual locations are the opposite. This mistake may be due to the wrong naming of chromosomes 1 and 26, or to a less probable ohnologous reciprocal translocation of the *noc*-containing segment.

Many methods have been published to quantify synteny, e.g., [41,42]. These bioinformatics tools are designed to analyze whole genomes in order to obtain large-scale synteny mapping based on small synteny blocks. However, the complexity of these tools is likely to yield widely divergent results, and better and simpler tools [43] are needed to achieve our objectives.

Based on synteny maps, we proposed two new indicators to quantify the degree of synteny conservation of neighboring *noc* genes by comparing two related species. The synteny index (SI) is useful for determining how many gene positions are conserved in a chromosomal fragment; we have found values higher than the threshold level. Moreover, the synteny conservation rate (SCR) reveals a conservation higher than the 50% of goldfish synteny with respect to zebrafish, common carp, and Prussian carp. Values obtained in these indicators match the phylogenetic tree topology (Figure 11C). In addition, these two numerical parameters may be used to confirm the correct assignment of a gene to its orthologous group. When these parameters are calculated in the ohnologous contrast (matrilineal vs. patrilineal chromosomes), the values drop considerably, indicating the absence of a real orthology between these ohnologous chromosomes (Supplementary Table S5).

The phylogenetic tree of protein sequences (Figure 12) supports and complements the conclusions of the synteny analysis. The tree in the present report follows the phylogenetic relationships of vertebrates in general, and teleosts in particular. It identifies the position or 3R teleostean genome duplication that generates the NOC-A and NOC-B branches. Furthermore, the 4Rc duplication segregates the cyprinine nocturnins in matrilineal and patrilineal paralogues.

We have included this phylogenetic information in the rationale of nocturnin nomenclature. We add to the zebrafish-based name a letter (a or b), which indicates the inclusion in patrilineal or matrilineal subgenomes, respectively. These criteria may be used for naming the rest of the cyprinine genes, and even extended to other allopolyploid groups. Then, the last letter of all gene names on a chromosome must be a or b, in reference to which subgenome the chromosome belongs. Cyprinidae underwent multiple and independent allopolyploidizations [44], but the same rationale can be extrapolated to other tetraploid groups, just by changing the last letter of gene names for another (different from a or b),

indicating different ancestral subgenomes. Thus, the allotetraploid clawed frog (*Xenopus laevis*) separates the genome in S and L subgenomes [45].

Current knowledge regarding the physiological actions of NOC in fish is very limited. The ubiquitous tissue expression of all *noc* paralogues in zebrafish and goldfish, in agreement with previous observations of the expression pattern of *noc-a* and *noc-b* [6], argues in favor of its involvement in many different physiological functions. It is important to consider that primers used for the quantification in the previous study in goldfish [6] were different from the ones used in the present report, and showed the total amount of both *noc-aa* and *noc-ab* summed together. Nevertheless, only one discrepancy is found in the two goldfish studies, which concerns the pituitary. The highest expression of *noc-aa* and *noc-ab* is shown in Figure 13, but a very low expression of total *noc-a* is described in the previous paper [6]. Present results in the brain allow us to conclude that the three paralogues are equally abundant in the forebrain, while the *noc-bb* paralogue is the most expressed in the mid- and hindbrain. This enriched expression of *noc-bb* in the optic tectum, vagal lobe and brainstem could be due to selective induction of this paralogue or repression of the other paralogues. Consistently, *noc-b* expression is three times that of *noc-a* in the brain of zebrafish, reinforcing the relevance of this paralogue. This wide distribution of *noc* in brain areas is also described in mice with different daily profiles of expression [7]. It is plausible to hypothesize a relationship between *noc*, as a clock-controlled gene, and the metabolic requirements of different brain areas [7]. Further studies exploring the rhythmicity of *noc* paralogues in specific brain areas of fish are needed.

All the *noc* paralogues are highly expressed in the retina of both goldfish and zebrafish, but *noc-ab* is more abundant in the subfamily Cyprininae, while *noc-b* is more abundant in Danioninae. This result could indicate an exchange of functional specialization of these paralogues since the divergence of these two cyprinid subfamilies. In zebrafish retina, the presence of *noc-a* transcripts is also detected (being ten times lower than *noc-b* transcripts), which could suggest some possible cooperative roles of all the *noc* for adequate retinal physiology. However, the only previous published study failed to detect the presence of *noc-a* transcripts in zebrafish retina [5], which could be explained by the different sensitivities of the experimental approaches in both studies. The cellular localization of *noc* expression in the retina is also controversial. In the clawed frog, *noc* expression is restricted to the photoreceptor cell layer [1]. Meanwhile, in mice, as in zebrafish, *noc* expression is high in photoreceptors, and lower (but clearly detected) in inner retina [46].

The peripheral distribution of *noc* paralogues is quite similar in zebrafish and goldfish, with some interesting exceptions. The broad expression of all paralogues throughout the entire digestive tract of goldfish and zebrafish supports the involvement of NOCs in several digestive processes in teleosts. Similar expression levels of all paralogues are found in the intestinal bulb of goldfish, and a brief fasting upregulated the expression of *noc-b*, but not *noc-a* paralogues [6]. This result and the highest expression of *noc-b* in the intestinal bulb of zebrafish (present results) could assign specialized functions for NOC-B in this region of digestive tract. The high amounts of *noc* transcripts in the hepatopancreas of goldfish and liver of zebrafish found in the present study agrees with results in mammals [7], and support their possible role in the hepatic lipid metabolism. Without physiological evidence and based only on their relative abundance, a more relevant involvement of NOC-A in teleosts, and particularly the NOC-AB paralogue in Cyprininae, could be suggested.

The adipose tissue shows the highest expression of all paralogues in goldfish but a moderate expression in zebrafish, with a significantly higher abundance of *noc-b* than *noc-a*. This result supports a general role for NOC in adipogenesis in Cyprininae, with NOC-B being the most relevant in Danioninae. A similar interspecific pattern is found in the anterior intestine, with enriched expression of both *noc-a* paralogues in goldfish but only of *noc-b* in zebrafish. A functional divergence of NOC proteins in the regulation of intestinal lipid absorption might be suggested, assumed of NOC-A in Cyprininae and of NOC-B in Danioninae. Finally, all the *noc* paralogues are present in the muscle, gonad, skin,

and head kidney in both goldfish and zebrafish, but there is no proposed function for NOC in these tissues.

4. Materials and Methods

4.1. Animal Maintenance

Goldfish (comet goldfish, *Carassius auratus auratus*) and zebrafish (*Danio rerio*, genotype Tübingen) with a body weight of 15.1 ± 2.5 g and 0.9 ± 0.2 g, respectively, were obtained from a local commercial supplier (Industrias Canarias del Acuario S.A., ICA, Madrid, Spain). Goldfish were kept in 60 L aquaria ($n = 6\text{--}7$ /aquarium) with filtered fresh water at 22 ± 1 °C, and zebrafish were maintained in 4 L racks ($n = 5$ /rack) at 27 ± 1 °C. All fishes were under a 12 h light–12 h darkness cycle (12L:12D photoperiod, lights on at 08:00 a.m., i.e., Zeitgeber Time 0, ZT0) with continuous aeration. A commercial flake diet (1.5% body weight, Sera Pond for goldfish, Tropifish ICA for zebrafish) was offered daily at 10:00 h (Zeitgeber time-2, ZT2) for at least three weeks before the assays. Fish were killed with an overdose of anesthesia (MS-222, Sigma-Aldrich, Madrid, Spain) at ZT4, and immediately sampled. All procedures were approved by the Animal Experimentation Committee of Complutense University and the Community of Madrid (PROEX 170.6/20), and according to the Guidelines of the European Union Council (2010/63/EU) and the Spanish Government (RD53/2013) for the use of animals in scientific research.

4.2. mRNA Sequencing of Goldfish Nocturnins

To obtain the full-length sequences, we used one 24 h fasted goldfish sacrificed at ZT4, and the whole hepatopancreas was collected.

Total mRNA was extracted using TRI Reagent (Merck, Madrid, Spain) following the manufacturer's instructions in mechanically homogenized hepatopancreas samples. Then, mRNA samples were treated with RQ1 RNase-Free DNase (Promega, Madison, WI, USA) according to a 4 µg mRNA/1 µL DNase ratio. Then, an aliquot of 1 µg of total RNA was retrotranscribed into cDNA in a 25 µL reaction volume using random primers (Invitrogen, Waltham, MA, USA), RNase inhibitor (Promega) and SuperScript II Reverse Transcriptase (Invitrogen). The reverse-transcription reaction consisted of 25 °C for 10 min, an extension of 50 min at 42 °C, and a denaturalization step at 70 °C for 15 min. The first strand cDNA fragments obtained were used as a template to amplify goldfish sequences with different primers (Merck), which were specifically designed (Supplementary Table S6).

The PCRs were performed in a 50 µL reaction volume containing 0.25 µL of HotStart DNA Polymerase recombinant 100 U, 10 µL PCR Buffer (5×, pH 8.5), 4 µL of MgCl₂ 25 mM (Promega), 1 µL of dNTP mixture 0.3 mM (Invitrogen), and 1 µL of each forward and reverse primer 10 µM and 2 µL of cDNA. The reaction conditions consisted of an initial denaturation at 94 °C for 3 min, followed by 40 cycles of 94 °C for 45 s, 57 °C or 60 °C for 30 s, and 72 °C for at least a 1 min/1 kb size sequence, with a final extension step at 72 °C for 10 min. For some reactions, touchdown PCRs were performed. PCR products were electrophoresed on a 0.5–2% agarose gel. Single bands for each PCR were purified using a GenElute™ Gel Extraction Kit (Sigma-Aldrich) and sequenced using the Sanger dideoxy method (Secugen, Madrid, Spain).

To obtain the consensus sequences for each paralogue, firstly each amplicon was validated as a *noc* sequence with the BLAST 2.14.0 program (<http://blast.ncbi.nlm.nih.gov/> accessed on 1 September 2023). All the *noc* sequences accepted as true were entered into the CAP3 program [47] (<http://doua.prabi.fr/software/cap3> accessed on 1 September 2023), and the resulting assembled sequence was accepted as consensus when a minimum of five independent sequenced amplicons were obtained for each bp position. Following this procedure, we detected two allelic variants in *noc-aa*.

To obtain the *noc* full-length sequences, and resolve the two allelic sequences for *noc-aa*, we used the 4-TOPO cloning kit for sequencing (K4575-02, Invitrogen). The full sequences of *noc* paralogues' splicing variants were amplified using specific primers for each. After performing the PCR, amplicons were purified. To insert the *noc* sequences into

the plasmids, 4 μL of amplified cDNA was incubated at 4 $^{\circ}\text{C}$ for at least 12 h with 1 μL of TOPO vector and 1 μL of salt solution from the cloning kit. Then, 5 μL of plasmid cDNA was added into a tube of chemically competent *E. coli* of the cloning kit and incubated for 30 min at 4 $^{\circ}\text{C}$, followed by a thermal shock of 45 s at 42 $^{\circ}\text{C}$ to facilitate the bacteria transformation. After that, 800 μL of SOC medium was pipetted into the *E. coli* tubes and incubated at 37 $^{\circ}\text{C}$ and 600 rpm for 1 h. Finally, different volumes of the primary culture were inoculated in Petri dishes with LB-agar, X-Gal and IPTG (Merck) for at least 24 h. Afterwards, transformed colonies were selected to perform a single subculture in LB medium at 37 $^{\circ}\text{C}$. Then, a new PCR using M13 primers (Merck) was performed to amplify the complete inserted sequences of plasmids, and their sizes were checked via electrophoresis in 1% agarose gel. Then, 24 h later, the plasmid cDNA was isolated from the subculture tubes (following the manufacturer's instructions) and sequenced using T3 and T7 primers (Secugen, Santa Clara, CA, USA). A minimum of 20 positive colonies were sequenced for each gene to validate the sequence obtained from cloning with the sequence obtained from PCR-CAP3 consensus.

4.3. Blast Analysis of mRNA SRA Libraries

To find the expression of different splicing variants of *noc* in the mRNA Sequence Read Archive (SRA) libraries, we selected SRA projects of *Carassius auratus* obtained from multiple organs, tissues, and developmental stages, available from NCBI databases. We designed a query sequence for each *noc* variant, consisting of 35 bp upstream and 35 bp downstream from the splicing point between exon 1 and exon 2. For finding matches, we used the Blastn 2.14.0 tool (<http://blast.ncbi.nlm.nih.gov/> accessed on 1 September 2023). For positive hit criteria, we accepted a minimum of query coverage of 66% and identity percentage of 96%. With these criteria, false positives are not counted, but SNPs (corresponding to allelic variants) are included.

4.4. In Silico Structural Analysis and Phylogenetic Tree Construction

An in silico structural analysis of the *noc* sequences obtained was carried out using online bioinformatics tools. The predictions of molecular weight and other basic properties of the protein were carried out using the ProtParam tool from ExPASy (<http://web.expasy.org/protparam/> accessed on 1 September 2023). The secondary structure prediction was performed using JPred4 (www.compbio.dundee.ac.uk/jpred/index_up.html accessed on 1 September 2023) [48]. JPred is a web server that uses the PSI-BLAST algorithm to generate secondary structures from the protein sequence. The tertiary structure was made using SWISS-MODEL (<https://swissmodel.expasy.org/> accessed on 1 September 2023) [49]. For searching MTS motifs, we used the MitoFates (<http://mitf.cbrc.jp/MitoFates/cgi-bin/top.cgi> accessed on 1 September 2023) [50], and TPpred 3.0 (<https://tppred3.biocomp.unibo.it> accessed on 1 October 2023) [51] tools. N-myristoylation probability is calculated with Myristoylator (<https://web.expasy.org/myristoylator/> accessed on 1 September 2023) [52], as assuming a high confidence in myristoylation when the score was higher than 0.85. Other molecular motifs were analyzed with the Eukaryotic Linear Motif (ELM) resource (<http://elm.eu.org/> accessed on 1 September 2023) [53]. Protein construction from WGS sequences was performed using the Genewise tool from EMBL-EBI (www.ebi.ac.uk/Tools/psa/genewise/ accessed on 1 September 2023) [54].

A phylogenetic analysis was performed by aligning the goldfish NOC amino acid sequences with those of other vertebrates retrieved from GenBank using the Clustal-X2.1 tool (www.clustal.org accessed on 1 September 2023) [55]. The evolutionary model and the phylogenetic tree were constructed with the help of MEGA 11 software [56]. After testing several phylogenetic methods, we opted for maximum likelihood (ML) with 1000 replicates for the bootstrap test. The representation was achieved using NjPlot 2.3 software (<http://pbil.univ-lyon1.fr/software/njplot.html> accessed on 1 September 2023) [57], considering *Homo sapiens*, *Xenopus tropicalis* and *Latimeria chalumnae* as outgroup species.

4.5. Synteny Analysis

The synteny analysis of *noc* was accomplished on the basis of the most recent assemblies of genomes published for the studied species (Supplementary Table S2). Two different websites (www.ncbi.nlm.nih.gov; www.ensembl.org accessed on 1 September 2023) published automatic (computer-based) annotation of predicted genes. These annotations have important differences, as they use different software resources, even though they are based on the same assembly project.

The graphic representation of synteny analysis was drawn using the automatically generated schemes from Genomicus 110.01 (www.genomicus.bio.ens.psl.eu/genomicus accessed on 1 September 2023) [58] as a draft copy, which uses genomic information from the Ensembl database. Furthermore, we manually curated the graphics by checking gene per gene in all the species considered, comparing the Genomicus synteny maps with the NCBI sequences, to confirm the loci analyzed and their conservation.

The nomenclature of the predicted genes was also reviewed, and the most suitable name was verified using the BLAST tool by comparing their sequences with the GenBank sequences that showed higher homology. Additionally, the ZFIN database (www.zfin.org accessed on 1 September 2023) [59] was also used for the gene nomenclature of *Danio rerio*.

For the final graphic representation, the scheme of the analyzed chromosomes was arranged in parallel, and each gene is represented by a pentagon whose sharp vertex indicates the sense of transcription on the chromosome (right: the coding strand is the direct one; left: the coding strand is the reverse complement). Orthologous genes are represented in the same color and are joined by straight lines to facilitate their location. In the center of the maps is the *Danio rerio* chromosome (with a white background), which has been used as a reference, and over that, there are the matrilineal chromosomes (with a pink background) in the following order: *Sinocyclocheilus grahami*, *Cyprinus carpio*, *Carassius gibelio*, and *Carassius auratus*. Towards the bottom, the patrilineal chromosomes (with a cyan background) are arranged in the corresponding symmetrical order: *Sinocyclocheilus grahami*, *Cyprinus carpio*, *Carassius gibelio*, and *Carassius auratus*.

To carry out a quantitative synteny analysis and the assessment of the synteny conservation for any gene cluster, two new parameters were determined: the synteny index (SI) and the synteny conservation rate (SCR) (Appendix A).

4.6. Tissue Expression of *noc* mRNA Using RT-qPCR

Samples of the telencephalon, diencephalon, optic tectum, hypothalamus, pituitary, cerebellum, vagal lobe, brainstem, retina, head kidney, gills, heart, esophagus, intestinal bulb, anterior intestine, middle intestine, posterior intestine, hepatopancreas, spleen, kidney, gonads, adipose tissue, skin, and muscle were collected from 6 goldfish at ZT4.

To analyze the mRNA distribution of *noc* in zebrafish, brain, retina, head kidney, gills, heart, esophagus, intestinal bulb, anterior intestine, posterior intestine, liver, pancreas, spleen, kidney, gonads, adipose tissue, skin, and muscle were collected from 5 fish at ZT4.

Total mRNA extraction and reverse-transcription of 2 µg of RNA were performed as described above. Then, real-time quantitative PCR (RT-qPCR) was performed using iTaq Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA). The specific primer sequences used for the target genes (Supplementary Table S6) were designed to amplify *noc-aa*, *noc-ab* and *noc-bb* specifically. The primers for goldfish were hand-made from full cds sequences obtained in the present work. Primers for the reference genes of goldfish (β -actin and *ef-1 α*) [60] and for zebrafish nocturnins, and the reference gene (β -actin) were previously described [5,61].

Genes were amplified using 1 µL of cDNA and 0.5 µL of each forward and reverse primer 10 µM in a final volume of 10 µL. Each PCR run included a standard curve of cDNA and a negative control (water). The qPCR cycling conditions consisted of a ramp of 95 °C for 30 s and 40 cycles of a two-step amplification program (95 °C for 5 s and 60 °C for 30 s), except in the case of goldfish *noc-bb*, wherein an additional step of 58 °C for 30 s was added to the amplification program. The melting curve was systematically monitored (with a

temperature gradient of 0.5 °C/5 s from 65 to 95 °C) at the end of each run to confirm the specificity of the amplification reaction. The $2^{-\Delta\Delta C_t}$ method was used to determine the relative mRNA expression [62], considering a the relative value of “1” for the brain in zebrafish and the hepatopancreas and diencephalon in the peripheral and neural regions of goldfish, respectively.

4.7. Statistics

Statistic analyses were carried out using SigmaPlot 12.0 (Systat Software Inc., San José, CA, USA). Data were checked for normality (Shapiro–Wilk test) and homoscedasticity (Levene’s test). Statistical differences in goldfish *noc-aa*, *noc-ab*, and *noc-ba*, or zebrafish *noc-a* and *noc-b* expression among different tissues were analyzed using a one-way ANOVA, followed by the post hoc Student–Newman–Keuls (SNK) multiple comparisons test. Significant differences were considered at $p < 0.05$.

5. Conclusions

This work presents the first full characterization of molecular structure of the different *nocturnin* paralogues in fish. The only human *noc* gene generates two (mitochondrial and cytoplasmic) isoforms of protein NOC through alternative translation initiation sites of a single mRNA, while in fish, the molecular diversity of NOC is achieved by whole genome duplication and by alternative splicing.

The *noc* gene of basal Osteichthyes has three splicing variants I, II and III, while in many genomes of Chondrichthyes, Osteichthyes, and Tetrapoda, only two variants (I and II) can be identified. From the four expected paralogues in goldfish, by direct mRNA sequencing, we obtained two isoforms for the *noc-a* subgroup (*noc-aa* and *noc-ab*) with two splice variants (I and II) each, and only one for the *noc-b* subgroup (*noc-bb*) with two transcripts (II and III). In genus *Carassius*, the other expected paralogue *noc-ba* is a pseudogene.

The splicing variant I of cyprinids and most fish contains an MTS motif that predestines this NOC variant to be a mitochondrial enzyme. Variant II has a conserved N-terminal glycine myristoylation motif followed by a hypervariable region with several serial serines that affect its myristoylation efficiency; it is species- and specimen-specific. Variant III has a very short coding exon 1c containing a DEG Nend UBRbox 2 motif, which means this variant is targeted for rapid ubiquitin-dependent proteasomal degradation. The core of the catalytic domain conserves the residues involved in Mg^{2+} -dependent phosphatase activity in all splicing variants.

From the synteny analysis and the phylogenetic tree of protein sequences, we proposed a classification and a long-lasting nomenclature of *noc* in goldfish extrapolated to allotetraploid Cyprininae, based in the localization of *noc* genes in the patrilineal or matrilineal subgenomes. We add to the zebrafish-based name a letter (a or b), indicating their inclusion in patrilineal or matrilineal subgenomes, respectively.

All the *noc* paralogues are ubiquitously expressed in goldfish and zebrafish, which support their possible involvement in several physiological functions. However, there are also interspecies differences regarding the relative abundance of the *noc* paralogues in several tissues.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25010054/s1>. References [63–67] are cited in Supplementary Materials.

Author Contributions: Conceptualization, D.M., M.J.D., A.I.V. and Á.L.A.-G.; methodology, D.M. and A.A.-G.; formal analysis, D.M. and Á.L.A.-G.; investigation, D.M., A.A.-G., M.J.D., A.I.V. and Á.L.A.-G.; writing—original draft preparation, D.M. and Á.L.A.-G.; writing—review and editing, M.J.D., A.I.V. and Á.L.A.-G.; visualization, D.M. and Á.L.A.-G.; supervision, M.J.D., A.I.V. and Á.L.A.-G.; project administration, M.J.D.; funding acquisition, M.J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Spanish Ministerio de Ciencia, Innovación y Universidades, PID2019-103969RBC32 and PID2022-136288OB-C32., D.M. is a predoctoral fellow from the Universidad Complutense de Madrid (CT42/18-CT43/18), and A.A.-G. is a predoctoral fellow from the Ministerio de Ciencia, Innovación y Universidades (BES-2017-081398).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, the regulations of the European Union Council (UE63/2010), and the Spanish Government (RD53/2013) for the use of animals in scientific proposals. It was approved by the Ethics Committee of the Complutense University of Madrid (O.H.-UCM-1310122019-2020 and by the Community of Madrid (PROEX 170.6/20 approval date: 21 July 2020).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are included in the article and the Supplementary Material.

Acknowledgments: The authors would like to thank the members of the Fish Neuroendocrinology Research Group (Universidad Complutense de Madrid) for their support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Synteny Quantitative Analysis:

- Gene in common (#). This is a gene present in two chromosomes, regardless of its position.
- Conserved Relative Position (*). The position of a gene is conserved when the upstream and downstream genes are the same on both chromosomes, and the sense of transcription is also conserved.

Quantitative Indicators:

Taking a gene as a reference, these indicators can quantify the synteny conservation between two whole homologous chromosomes containing it, or as is more common, between two chromosome segments, indicating how many genes upstream and downstream of the reference gene are used.

1. The Synteny Index (SI) determines the ratio between conserved and non-conserved relative positions in segments of two orthologous chromosomes. A synteny index higher than 1 indicates more conserved relative positions than non-conserved ones. In the case of high conservation (in which the non-conserved relative positions tend to be 0), the SI tends to be infinite.

$$\text{Synteny Index (SI)}_{(no.upstream\ genes, no.downstream\ genes)} = \frac{no.\ conserved\ relative\ positions}{(no.\ conserved\ genes - no.\ conserved\ relative\ positions)}$$

2. The Synteny Conservation Rate (SCR) follows the Laplace law for probability. This index expresses the percentage of conservation of relative positions with respect to the total number of genes in common in two orthologous chromosomes.

$$\text{Synteny Conservation Rate (SCR)}_{(no.upstream\ genes, no.downstream\ genes)} = \frac{no.\ conserved\ relative\ positions}{no.\ genes\ in\ common} \times 100$$

The SCR will be 100% only when two entire chromosomes are compared, considering that the genes adjacent to the telomere might only conserve one neighbor locus, and be counted as a “conserved relative position”.

Even though chromosome blocks with a perfect match are analyzed, when SCR is applied to a segment, is not possible to reach the 100%, because the gene at the end of a chromosomal segment cannot be considered as a “conserved relative position”. The

theoretical maximum synteny conservation rate (mSCR) in two chromosome segments is calculated using the following formula:

$$\text{maximum Synteny Conservation Rate (mSCR)} = \frac{\text{no. genes in common} - \text{no. end nontelomeric genes}}{\text{no. genes in common}} \times 100$$

(no.upstream genes, no.downstream genes)

Example of Synteny Calculation of F gene:

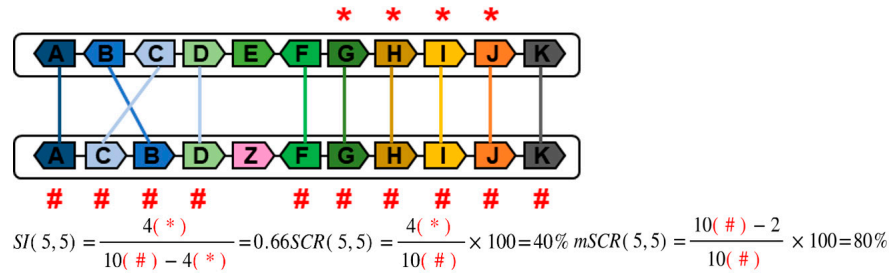


Figure A1. Schematic representation of two orthologous chromosome segments. White boxes represent the chromosomes. Genes are represented with pentagons whose sharp vertex indicates the transcription sense. Same colors and letters indicate same genes. *, genes with conserved relative position. #, genes in common.

Special cases to be considered in the identification of conserved relative positions:

- Tandem Duplications

If the tandem duplication is maintained in the two orthologous chromosome segments analyzed, we consider that the genes and the relative position are conserved (Figure A2A). If the tandem duplication is only in one of the two orthologous chromosome segments, by parsimony, we assume that this gene is conserved from the ancestral chromosome (with only one copy), and that it has undergone tandem duplication in the other chromosome (with two copies).

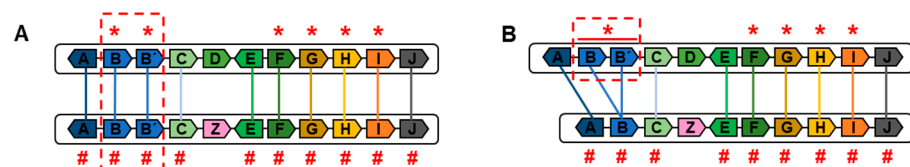


Figure A2. Schematic representation of tandem duplication in two (A) or one (B) orthologous chromosome segments. White boxes represent the chromosomes. Genes are represented with pentagons whose sharp vertex indicates the transcription sense. Similar colors and letters indicate the same genes. Red boxes indicate tandem duplication. * refers to genes with a conserved relative position between both chromosome segments. # refers to genes in common on both chromosome segments.

- Inversions

If only one gene is inverted (single inversion), its relative position is non-conserved considering the change in the transcription sense. The relative position of the adjacent genes to the single inversion is not conserved either (Figure A3A).

When a gene cluster is inverted (with several genes), only the relative position of the extreme genes of the cluster are non-conserved. Genes located between the extremes of the cluster have conserved the relative position, considering that the upstream and downstream positions and the sense of transcription are the same in both orthologous chromosome segments (Figure A3B).

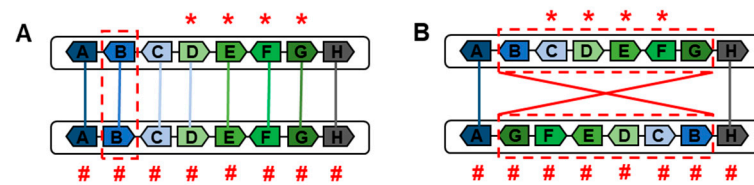


Figure A3. Schematic representation of a single (A) and a gene cluster (B) inversion in two orthologous chromosome segments. White boxes represent the chromosomes. Genes are represented with pentagons whose sharp vertex indicates the transcription sense. Similar colors and letters indicate the same genes. Red boxes indicate the gene and inverted gene cluster. * refers to genes with a conserved relative position between both chromosome segments. # refers to genes in common on both chromosome segments.

- Translocations

With this mutation, it is considered that only the relative position of the extreme genes of the translocation is non-conserved. The rest of the genes included in the translocated segment conserve their relative positions because the up and downstream genes maintain their loci independently of the chromosome wherein they are located (Figure A4).

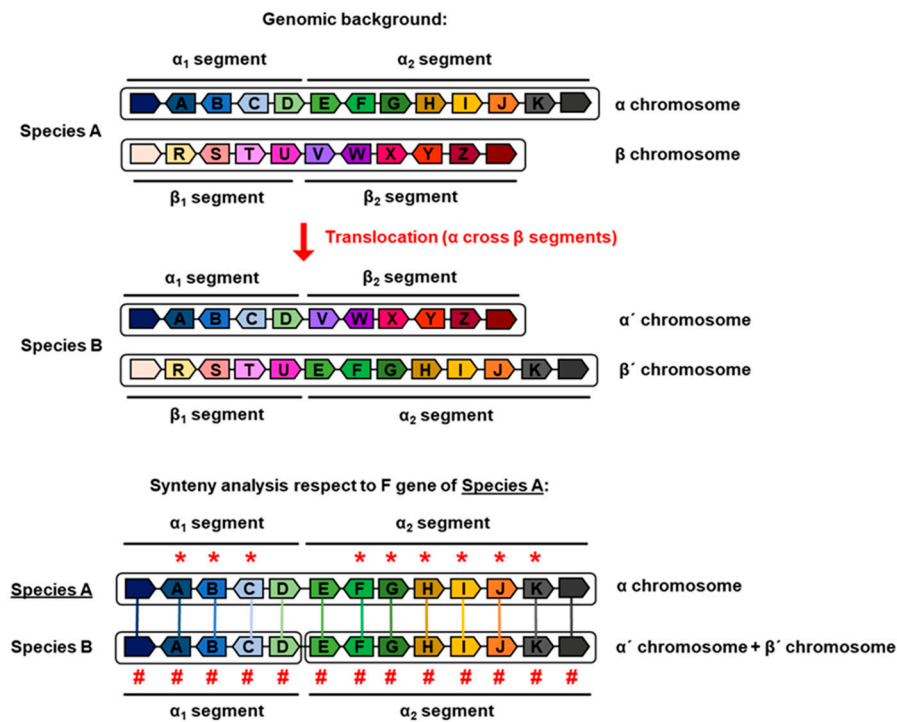


Figure A4. Schematic representation of a synteny analysis in a translocation of two orthologous chromosomes. White boxes indicate the chromosomes. Genes are represented with pentagons whose sharp vertex indicates the transcription sense. Similar colors and/or letters indicate the same genes. The red arrow indicates the translocation point. * refers to genes with a conserved relative position between both chromosome segments. # refers to genes in common on both chromosome segments.

References

- Green, C.B.; Besharse, J.C. Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14884–14888. [CrossRef] [PubMed]
- Müller, W.E.G.; Wang, X.; Grebenjuk, V.A.; Korzhev, M.; Wiens, M.; Schlossmacher, U.; Schröder, H.C. Nocturnin in the demosponge *Suberites domuncula*: A potential circadian clock protein controlling glycogenin synthesis in sponges. *Biochem. J.* **2012**, *448*, 233–242. [CrossRef] [PubMed]
- Chang, V.; Meuti, M.E. Circadian transcription factors differentially regulate features of the adult overwintering diapause in the Northern house mosquito, *Culex pipiens*. *Insect Biochem. Mol. Biol.* **2020**, *121*, 103365. [CrossRef] [PubMed]

4. Grönke, S.; Bickmeyer, I.; Wunderlich, R.; Jäckle, H.; Kühnlein, R.P. Curled encodes the *Drosophila* homolog of the vertebrate circadian deadenylase nocturnin. *Genetics* **2009**, *183*, 219–232. [[CrossRef](#)] [[PubMed](#)]
5. Yang, X.; Fu, J.; Wei, X. Expression patterns of zebrafish nocturnin genes and the transcriptional activity of the frog nocturnin promoter in zebrafish rod photoreceptors. *Mol. Vis.* **2017**, *23*, 1039–1047.
6. Blanco, A.M.; Gómez-Boronat, M.; Madera, D.; Valenciano, A.I.; Alonso-Gómez, A.L.; Delgado, M.J. First evidence of nocturnin in fish: Two isoforms in goldfish differentially regulated by feeding. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **2018**, *314*, R304–R312. [[CrossRef](#)]
7. Wang, Y.; Osterbur, D.L.; Megaw, P.L.; Tosini, G.; Fukuhara, C.; Green, C.B.; Besharse, J.C. Rhythmic expression of Nocturnin mRNA in multiple tissues of the mouse. *BMC Dev. Biol.* **2001**, *1*, 9–15. [[CrossRef](#)]
8. Li, R.; Yue, J.; Zhang, Y.; Zhou, L.; Hao, W.; Yuan, J.; Qiang, B.; Ding, J.M.; Peng, X.; Cao, J.M. CLOCK/BMAL1 regulates human nocturnin transcription through binding to the E-box of nocturnin promoter. *Mol. Cell. Biochem.* **2008**, *317*, 169–177. [[CrossRef](#)]
9. Curran, K.L.; LaRue, S.; Bronson, B.; Solis, J.; Trow, A.; Sarver, N.; Zhu, H. Circadian genes are expressed during early development in *Xenopus laevis*. *PLoS ONE* **2008**, *3*, e2749. [[CrossRef](#)]
10. Bracci, M.; Copertaro, A.; Ciarapica, V.; Barbaresi, M.; Esposito, S.; Albanesi, A.; Valentino, M.; Ledda, C.; Rapisarda, V.; Santarelli, L. Nocturnin gene diurnal variation in healthy volunteers and expression levels in shift workers. *Biomed. Res. Int.* **2019**, *31*, 7582734. [[CrossRef](#)]
11. Liu, X.; Green, C.B. Circadian regulation of nocturnin transcription by phosphorylated CREB in *Xenopus* retinal photoreceptor cells. *Mol. Cell. Biol.* **2002**, *22*, 7501–7511. [[CrossRef](#)] [[PubMed](#)]
12. Baggs, J.E.; Green, C.B. Nocturnin, a deadenylase in *Xenopus laevis* retina: A mechanism for postranscriptional control of circadian-related mRNA. *Curr. Biol.* **2003**, *13*, 189–198. [[CrossRef](#)] [[PubMed](#)]
13. Kojima, S.; Gendreau, K.L.; Sher-Chen, E.L.; Gao, P.; Green, C.B. Changes in poly(A) tail length dynamics from the loss of the circadian deadenylase Nocturnin. *Sci. Rep.* **2015**, *5*, 17059. [[CrossRef](#)] [[PubMed](#)]
14. Stubblefield, J.J.; Terrien, J.; Green, C.B. Nocturnin: At the crossroads of clocks and metabolism. *Trends Endocrinol. Metab.* **2012**, *23*, 326–333. [[CrossRef](#)] [[PubMed](#)]
15. Estrella, M.A.; Du, J.; Chen, L.; Rath, S.; Prangley, E.; Chitrakar, A.; Aoki, T.; Schedl, P.; Rabinowitz, J.; Korennykh, A. The metabolites NADP⁺ and NADPH are the targets of the circadian protein Nocturnin (Curled). *Nat. Comm.* **2019**, *10*, 2367. [[CrossRef](#)] [[PubMed](#)]
16. Green, C.B.; Douris, N.; Kojima, S.; Strayer, C.A.; Fogerty, J.; Lourim, D.; Keller, S.R.; Besharse, J.C. Loss of Nocturnin, a circadian deadenylase, confers resistance to hepatic steatosis and diet induced obesity. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9888–9893. [[CrossRef](#)] [[PubMed](#)]
17. Douris, N.; Kojima, S.; Pan, X.; Lerch-Gaggl, A.F.; Duong, S.Q.; Hussain, M.; Green, C.B. Nocturnin regulates circadian trafficking of dietary lipid in intestinal enterocytes. *Curr. Biol.* **2011**, *21*, 1347–1355. [[CrossRef](#)] [[PubMed](#)]
18. Douris, N.; Green, C.B. NOC out the fat: A short review of the circadian deadenylase Nocturnin. *Ann. Med.* **2008**, *40*, 622–626. [[CrossRef](#)]
19. Stubblefield, J.J.; Gao, P.; Kilaru, G.; Mukadam, B.; Terrien, J.; Green, C.B. Temporal control of metabolic amplitude by nocturnin. *Cell Rep.* **2018**, *22*, 1225–1235. [[CrossRef](#)]
20. Kawai, M.; Green, C.B.; Lecka-Czernik, B.; Douris, N.; Gilbert, M.R.; Kojima, S.; Ackert-Bicknell, C.; Garg, N.; Horowitz, M.C.; Adamo, M.L.; et al. A circadian-regulated gene, Nocturnin, promotes adipogenesis by stimulating PPAR-gamma nuclear translocation. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10508–10513. [[CrossRef](#)]
21. Kawai, M.; Delany, A.M.; Green, C.B.; Adamo, M.L.; Rosen, C.J. Nocturnin suppresses igf1 expression in bone by targeting the 3' untranslated region of igf1 mRNA. *Endocrinology* **2010**, *151*, 4861–4870. [[CrossRef](#)] [[PubMed](#)]
22. Nishikawa, S.; Hatanaka, Y.; Tokoro, M.; Shin, S.-W.; Shimizu, N.; Nishihara, T.; Kato, R.; Takemoto, A.; Amano, T.; Anzai, M.; et al. Functional analysis of nocturnin, a circadian deadenylase, at maternal-to-zygotic transition in mice. *J. Reprod. Dev.* **2013**, *59*, 258–265. [[CrossRef](#)] [[PubMed](#)]
23. Curran, K.L.; Allen, L.; Porter, B.B.; Dodge, J.; Lope, C.; Willadsen, G.; Fisher, R.; Johnson, N.; Campbell, E.; VonBergen, B.; et al. Circadian genes, xBmal1 and xNocturnin, modulate the timing and differentiation of somites in *Xenopus laevis*. *PLoS ONE* **2014**, *9*, e108266. [[CrossRef](#)] [[PubMed](#)]
24. Meyer, A.; Van de Peer, Y. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *BioEssays* **2005**, *27*, 937–945. [[CrossRef](#)] [[PubMed](#)]
25. Macqueen, D.J.; Johnston, I.A. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. R. Soc. B Biol. Sci.* **2014**, *281*, 20132881. [[CrossRef](#)] [[PubMed](#)]
26. Yang, L.; Sado, T.; Vincent Hirt, M.; Pasco-Viel, E.; Arunachalam, M.; Li, J.; Wang, X.; Freyhof, J.; Saitoh, K.; Simons, A.M.; et al. Phylogeny and polyploidy: Resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* **2015**, *85*, 97–116. [[CrossRef](#)] [[PubMed](#)]
27. Abshire, E.T.; Chasseur, J.; Bohn, J.A.; Del Rizzo, P.A.; Freddolino, P.L.; Goldstrohm, A.C.; Trievel, R.C. The structure of human Nocturnin reveals a conserved ribonuclease domain that represses target transcript translation and abundance in cells. *Nucleic Acids Res.* **2018**, *46*, 6257–6270. [[CrossRef](#)]
28. Glasauer, S.M.; Neuhauss, S.C. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics.* **2014**, *289*, 1045–1060. [[CrossRef](#)]

29. Luo, J.; Chai, J.; Wen, Y.; Tao, M.; Lin, G.; Liu, X.; Ren, L.; Chen, Z.; Wu, S.; Li, S.; et al. From asymmetrical to balanced genomic diversification during rediploidization: Subgenomic evolution in allotetraploid fish. *Sci. Adv.* **2020**, *6*, eaaz7677. [[CrossRef](#)]
30. Laothamatas, I.; Gao, P.; Wickramaratne, A.; Quintanilla, C.G.; Dino, A.; Khan, C.A.; Liou, J.; Green, C.B. Spatiotemporal regulation of NADP(H) phosphatase Nocturnin and its role in oxidative stress response. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 993–999. [[CrossRef](#)]
31. Moe, S.E.; Sorbo, J.G.; Holen, T. Huntingtin triplet-repeat locus is stable under long-term Fen1 knockdown in human cells. *J. Neurosci. Methods* **2008**, *171*, 233–238. [[CrossRef](#)] [[PubMed](#)]
32. Tabolacci, E.; Nobile, V.; Pucci, C.; Chiurazzi, P. Mechanisms of the FMR1 repeat instability: How does the CGG sequence expand? *Int. J. Mol. Sci.* **2022**, *23*, 5425. [[CrossRef](#)] [[PubMed](#)]
33. Abshire, E.T.; Hughes, K.L.; Diao, R.; Pearce, S.; Gopalakrishna, S.; Trievel, R.C.; Rorbach, J.; Freddolino, P.L.; Goldstrohm, A.C. Differential processing and localization of human Nocturnin controls metabolism of mRNA and nicotinamide adenine dinucleotide cofactors. *J. Biol. Chem.* **2020**, *295*, 15112–15133. [[CrossRef](#)] [[PubMed](#)]
34. Poveda-Huertes, D.; Mulica, P.; Vögtle, F.N. The versatility of the mitochondrial presequence processing machinery: Cleavage, quality control and turnover. *Cell Tissue Res.* **2017**, *367*, 73–81. [[CrossRef](#)]
35. Sriram, S.M.; Kim, B.Y.; Kwon, Y.T. The N-end rule pathway: Emerging functions and molecular principles of substrate recognition. *Nat. Rev. Mol. Cell. Biol.* **2011**, *12*, 735–747. [[CrossRef](#)]
36. Alonso-Gómez, A.L.; Iuvone, P.M. Melatonin biosynthesis in cultured chick retinal photoreceptor cells: Calcium and cyclic AMP protect serotonin N-acetyltransferase from inactivation in cycloheximide-treated cells. *J. Neurochem.* **1995**, *65*, 1054–1060. [[CrossRef](#)]
37. Clark, L.B.; Viswanathan, P.; Quigley, G.; Chiang, Y.C.; McMahan, J.S.; Yao, G.; Chen, J.; Nelsbach, A.; Denis, C.L. Systematic mutagenesis of the leucine-rich repeat (LRR) domain of CCR4 reveals specific sites for binding to CAF1 and a separate critical role for the LRR in CCR4 deadenylase activity. *J. Biol. Chem.* **2004**, *279*, 13616–13623. [[CrossRef](#)]
38. Dupressoir, A.; Morel, A.P.; Barbot, W.; Loireau, M.P.; Corbo, L.; Heidmann, T. Identification of four families of yCCR4- and Mg²⁺-dependent endonuclease-related proteins in higher eukaryotes, and characterization of orthologs of yCCR4 with a conserved leucine-rich repeat essential for hCAF1/hPOP2 binding. *BMC Genom.* **2001**, *2*, 9. [[CrossRef](#)]
39. Godwin, A.R.; Kojima, S.; Green, C.B.; Wilusz, J. Kiss your tail goodbye: The role of PARN, Nocturnin, and Angel deadenylases in mRNA biology. *Biochim. Biophys. Acta* **2013**, *1829*, 571–579. [[CrossRef](#)]
40. Landschulz, W.H.; Johnson, P.F.; McKnight, S.L. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **1988**, *240*, 1759–1764. [[CrossRef](#)]
41. Sinha, A.U.; Meller, J. Cinteny: Flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. *BMC Bioinform.* **2007**, *8*, 82. [[CrossRef](#)] [[PubMed](#)]
42. Ali, R.H.; Muhammad, S.A.; Arvestad, L. GenFamClust: An accurate, synteny-aware and reliable homology inference algorithm. *BMC Evol. Biol.* **2016**, *16*, 120. [[CrossRef](#)] [[PubMed](#)]
43. Ghiurcuta, C.G.; Moret, B.M. Evaluating synteny for improved comparative studies. *Bioinformatics* **2014**, *30*, i9–i18. [[CrossRef](#)] [[PubMed](#)]
44. Yang, L.; Naylor, G.J.P.; Mayden, R.L. Deciphering reticulate evolution of the largest group of polyploid vertebrates, the subfamily cyprininae (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* **2022**, *166*, 107323. [[CrossRef](#)] [[PubMed](#)]
45. Session, A.M.; Uno, Y.; Kwon, T.; Chapman, J.A.; Toyoda, A.; Takahashi, S.; Fukui, A.; Hikosaka, A.; Suzuki, A.; Kondo, M.; et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **2016**, *538*, 336–343. [[CrossRef](#)] [[PubMed](#)]
46. Barbot, W.; Wasowicz, M.; Dupressoir, A.; Versaux-Botteri, C.; Heidmann, T. A murine gene with circadian expression revealed by transposon insertion: Self-sustained rhythmicity in the liver and the photoreceptors. *Biochim. Biophys. Acta* **2002**, *1576*, 81–91. [[CrossRef](#)] [[PubMed](#)]
47. Huang, X.; Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **1999**, *9*, 868–877. [[CrossRef](#)]
48. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **2015**, *43*, W389–W394. [[CrossRef](#)]
49. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)]
50. Fukasawa, Y.; Tsuji, J.; Fu, S.C.; Tomii, K.; Horton, P.; Imai, K. MitoFates: Improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteom.* **2015**, *14*, 1113–1126. [[CrossRef](#)]
51. Savojardo, C.; Martelli, P.L.; Fariselli, P.; Casadio, R. TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins. *Bioinformatics* **2015**, *31*, 3269–3275. [[CrossRef](#)] [[PubMed](#)]
52. Bologna, G.; Yvon, C.; Duvaud, S.; Veuthey, A.L. N-Terminal myristoylation predictions by ensembles of neural networks. *Proteomics* **2004**, *4*, 1626–1632. [[CrossRef](#)] [[PubMed](#)]
53. Kumar, M.; Michael, S.; Alvarado-Valverde, J.; Mészáros, B.; Sámano-Sánchez, H.; Zeke, A.; Dobson, L.; Lazar, T.; Örd, M.; Nagpal, A.; et al. The eukaryotic linear motif resource: 2022 release. *Nucleic Acids Res.* **2022**, *50*, D497–D508. [[CrossRef](#)] [[PubMed](#)]
54. Madeira, F.; Pearce, M.; Tivey, A.R.N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276–W279. [[CrossRef](#)]

55. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
56. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [[CrossRef](#)]
57. Perrière, G.; Gouy, M. WWW-Query: An on-line retrieval system for biological sequence banks. *Biochimie* **1996**, *78*, 364–369. [[CrossRef](#)]
58. Nguyen, N.T.T.; Vincens, P.; Dufayard, J.F.; Roest Crollius, H.; Louis, A. Genomicus in 2022: Comparative tools for thousands of genomes and reconstructed ancestors. *Nucleic Acids Res* **2022**, *50*, D1025–D1031. [[CrossRef](#)]
59. Bradford, Y.M.; Van Slyke, C.E.; Ruzicka, L.; Singer, A.; Eagle, A.; Fashena, D.; Howe, D.G.; Frazer, K.; Martin, R.; Paddock, H.; et al. Zebrafish information network, the knowledgebase for *Danio rerio* research. *Genetics* **2022**, *220*, iyac016. [[CrossRef](#)]
60. Sánchez-Bretaña, A.; Blanco, A.M.; Unniappan, S.; Kah, O.; Gueguen, M.M.; Bertucci, J.I.; Alonso-Gómez, A.L.; Valenciano, A.I.; Isorna, E.; Delgado, M.J. In Situ localization and rhythmic expression of ghrelin and ghs-r1 ghrelin receptor in the brain and gastrointestinal tract of goldfish (*Carassius auratus*). *PLoS ONE* **2015**, *10*, e0141043. [[CrossRef](#)]
61. Zhang, S.; Xu, M.; Huang, J.; Tang, L.; Zhang, Y.; Wu, J.; Lin, S.; Wang, H. Heme acts through the Bach1b/Nrf2a-MafK pathway to regulate exocrine peptidase precursor genes in porphyric zebrafish. *Dis. Model. Mech.* **2014**, *7*, 837–845. [[CrossRef](#)] [[PubMed](#)]
62. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]
63. Howe, K.; Clark, M.D.; Torroja, C.F.; Torrance, J.; Berthelot, C.; Muffato, M.; Collins, J.E.; Humphray, S.; McLaren, K.; Matthews, L.; et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **2013**, *496*, 498–503. [[CrossRef](#)] [[PubMed](#)]
64. Chen, Z.; Omori, Y.; Koren, S.; Shirokiya, T.; Kuroda, T.; Miyamoto, A.; Wada, H.; Fujiyama, A.; Toyoda, A.; Zhang, S.; et al. De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. *Sci. Adv.* **2019**, *5*, eaav0547. [[CrossRef](#)]
65. Kuhl, H.; Du, K.; Schartl, M.; Kalous, L.; Stöck, M.; Lamatsch, D.K. Equilibrated evolution of the mixed auto-/allopolyploid haplotype-resolved genome of the invasive hexaploid Prussian carp. *Nat. Commun.* **2022**, *13*, 4092. [[CrossRef](#)]
66. Xu, P.; Xu, J.; Liu, G.; Chen, L.; Zhou, Z.; Peng, W.; Jiang, Y.; Zhao, Z.; Jia, Z.; Sun, Y.; et al. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nat. Commun.* **2019**, *10*, 4625. [[CrossRef](#)]
67. Yang, J.; Chen, X.; Bai, J.; Fang, D.; Qiu, Y.; Jiang, W.; Yuan, H.; Bian, C.; Lu, J.; He, S.; et al. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. *BMC Biol.* **2016**, *14*, 1. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.