The fatty acid elongation genes *elovl4a* and *elovl4b* are present and functional in the genome of tambaqui (*Colossoma macropomum*)

Renato B. Ferraz^{a,b}, André M. Machado^a, Juan C. Navarro^c, Isabel Cunha^a, Rodrigo Ozório^a, Ana L. Salaro^d, L. Filipe C. Castro^{a,e†} and Óscar Monroig^{c†}

^aCIIMAR – Interdisciplinary Centre of Marine and Environmental Research – University of Porto, Portugal

^bICBAS - Institute of Biomedical Sciences Abel Salazar, University of Porto, Portugal

^cInstituto de Acuicultura Torre de la Sal, Consejo Superior de Investigaciones Científicas (IATS-CSIC), Spain

^dDepartment of Animal Biology, Federal University of Viçosa, Brazil

^eDepartment of Biology, Faculty of Sciences, University of Porto, Portugal

† Corresponding Authors: filipe.castro@ciimar.up.pt; oscar.monroig@csic.es

Address:

^a CIIMAR, University of Porto - Av. General Norton De Matos S/N, 4450-208 Matosinhos, Portugal

^b ICBAS, University of Porto - Rua de Jorge Viterbo Ferreira, 228, 4050-313, Porto, Portugal

^c Instituto de Acuicultura Torre de la Sal, Consejo Superior de Investigaciones Científicas (IATS-CSIC), 12595 Ribera de Cabanes, Castellón, Spain

^d Federal University of Viçosa - Avenida Peter Henry Rolfs, s/n - Campus Universitário, Viçosa - MG, 36570-900, Brazil

^e Department of Biology, Faculty of Sciences, University of Porto - Rua Campo Alegre s/n, 4169-007, Portugal

25 **Highlights**

- 26 • Elov14a and Elov14b were identified and functionally characterized from
27 tambaqui.
- 28 • A low coverage draft genome assembly from tambaqui was produced.
- 29 • Tambaqui Elov14a and Elov14b elongated products of up to C₃₆.
- 30 • Tambaqui Elov14 enzymes contribute to DHA biosynthesis.

31

32

Abstract

Long-chain (C₂₀₋₂₄) polyunsaturated fatty acids (LC-PUFA) are physiologically important nutrients for vertebrates including fish. Previous studies have addressed the metabolism of LC-PUFA in the Amazonian teleost tambaqui (*Colossoma macropomum*), an emerging species in Brazilian aquaculture, showing that all the desaturase and elongase activities required to convert C₁₈ polyunsaturated fatty acids (PUFA) into LC-PUFA are present in tambaqui. Yet, elongation of very long-chain fatty acid 4 (Elovl4) proteins, which participate in the biosynthesis of very long-chain (>C₂₄) saturated fatty acids (VLC-SFA) and very long-chain polyunsaturated fatty acids (VLC-PUFA), had not been characterized in this species. Here, we investigate the repertoire and function of two Elovl4 in tambaqui. Furthermore, we present the first draft genome assembly from tambaqui, and demonstrated the usefulness of this resource in nutritional physiology studies by isolating one of the tambaqui *elovl4* genes. Our results showed that, similarly to other teleost species, two *elovl4* gene paralogs termed as *elovl4a* and *elovl4b*, are present in tambaqui. Tambaqui *elovl4a* and *elovl4b* have open reading frames (ORF) of 948 and 912 base pairs, encoding putative proteins of 315 and 303 amino acids, respectively. Functional characterization in yeast showed that both Elovl4 enzymes have activity towards all the PUFA substrates assayed (18:3n-3, 18:2n-6, 18:4n-3, 18:3n-6, 20:5n-3, 20:4n-6, 22:5n-3, 22:4n-6 and 22:6n-3), producing elongated products of up to C₃₆. Moreover, both Elovl4 were able to elongate 22:5n-3 to 24:5n-3, a key elongation step required for the synthesis of docosahexaenoic acid via the Sprecher pathway.

Keywords

Biosynthesis; *Colossoma macropomum*; Elovl4; genome; very long-chain fatty acids

1. Introduction

Long-chain (C₂₀₋₂₄) polyunsaturated fatty acids (LC-PUFA) are physiologically important nutrients for vertebrates and especially, the so-called “omega-3” (or n-3) LC-PUFA play important roles in human health (Calder, 2014; Campoy et al., 2012; Delgado-Lista et al., 2012; Innis, 2007). Oily fish and seafood in general are primary sources of n-3 LC-PUFA in the human diet. LC-PUFA are commonly regarded as essential nutrients for many vertebrates since they cannot be biosynthesized at relevant physiological rates and, consequently, LC-PUFA must be supplied in the diet to prevent deficiency symptoms (Castro et al., 2016; Tocher, 2015). In some teleosts, particularly freshwater species (Ishikawa et al., 2019), dietary shorter chain C₁₈ polyunsaturated fatty acids (PUFA), such as linoleic (LA; 18:2n-6) and α -linolenic acids (ALA; 18:3n-3), can satisfy the essential fatty acid requirements since they can convert them into LC-PUFA (Castro et al. 2016; Monroig et al. 2018). Conversions of C₁₈ PUFA into C₂₀₋₂₄ LC-PUFA are mediated by members of two families of enzymes named fatty acyl desaturases (Fads) and elongation of very long-chain fatty acid (Elovl) proteins (Castro et al., 2016; Monroig and Kabeya, 2018; Tocher, 2015). Elovl are rate-limiting enzymes enabling the addition of two-carbon atoms to pre-existing fatty acid (FA) substrates (Guillou et al., 2010; Jakobsson et al., 2006). While a recent study has reported on the putative roles of a novel Elovl present, Elovl8, in PUFA elongation in teleosts (Li et al., 2020), robust evidence supports that mainly three members of the Elovl family, namely Elovl 2, 4 and 5, have been shown to participate in the biosynthesis of LC-PUFA in vertebrates (Jakobsson et al. 2006; Guillou et al., 2010; Castro et al 2016). More specifically, Elovl5 exhibits high activity toward C₁₆₋₂₀ PUFA substrates, while Elovl2 shows high activity toward C₂₀₋₂₂ PUFA (Jakobsson et al., 2006; Monroig et al., 2013). Elovl4 is involved in the biosynthesis of very long-chain (>C₂₄) FA including saturated (VLC-SFA) and polyunsaturated (VLC-PUFA) compounds (Deak et al., 2019).

Teleost fish possess two *elovl4* genes termed *elovl4a* and *elovl4b* (Monroig et al., 2010). Teleost *elovl4* genes have been so far isolated and characterized in zebrafish (*Danio rerio*) (Monroig et al., 2010), Atlantic salmon (*Salmo salar*) (Carmona-Antonanzas et al., 2011), cobia (*Rachycentron canadum*) (Monroig et al., 2011), rabbitfish (*Siganus canaliculatus*) (Monroig et al., 2012), Nibe croaker (*Nibea mitsukurii*) (Kabeya et al.,

2015), orange-spotted grouper (*Epinephelus coioides*) (Li et al., 2017a), loach (*Misgurnus anguillicaudatus*) (Yan et al., 2017), African catfish (*Clarias gariepinus*) (Obloh et al., 2017a), large yellow croaker (*Larimichthys crocea*) (Li et al., 2017b), black seabream (*Acanthopagrus schlegelii*) (Jin et al., 2017), rainbow trout (*Oncorhynchus mykiss*) (Zhao et al., 2019), and Atlantic bluefin tuna (*Thunnus thynnus*) (Betancor et al., 2020). Tissue distribution of *elovl4* mRNA suggests that this gene has a widespread expression pattern, with *elovl4a* being mostly expressed in brain tissues (brain and pituitary) (Monroig et al., 2010; Obloh et al., 2017a), and *elovl4b* in retina and gonads (Monroig et al., 2010; Obloh et al., 2017a). Teleost Elov4 functions were first characterized in *D. rerio*, confirming that Elov4a had the ability to elongate saturated fatty acids to produce VLC-SFA, whereas only Elov4b appeared to have a role in the biosynthesis of VLC-PUFA (Monroig et al., 2010). More recently however, it has been shown that both Elov4a and Elov4b from *Clarias gariepinus* and *Acanthopagrus schlegelii* have the ability to biosynthesize VLC-PUFA (Jin et al., 2017; Obloh et al., 2017a). Along with their role in VLC-PUFA biosynthesis, certain teleost Elov4 characterized to date have also shown ability to elongate 22:5n-3 to 24:5n-3 for biosynthesis of DHA through the so-called “Sprecher pathway” (Sprecher, 2000), a metabolic route operated in teleosts (Buzzi et al., 1996; Buzzi et al., 1997; Obloh et al., 2017b; Tocher et al., 2003).

Genome information is fundamental to decipher multiple aspects of a species biology. Various teleost species already have the genome assembled and annotated, in many cases for evolutionary study, mainly because during the early evolution of teleosts there was a genomic duplication (TGD, teleost genome duplication) that greatly increased the number of genes (Berthelot et al., 2014). Furthermore, aquaculture has also taken advantage of genomes available from commercially important species such as carp (*Cyprinus carpio*) (Xu et al., 2014), Atlantic salmon (*S. salar*) (Lien et al., 2016), rainbow trout (*O. mykiss*) (Palti et al., 2012), European sea bass (*Dicentrarchus labrax*) (Tine et al., 2014), European sardine (*Sardina pilchardus*) (Louro et al., 2019; Machado et al., 2018) and Nile tilapia (*Oreochromis niloticus*) (Conte et al., 2017). Similarly, in Brazil, sequencing projects of neotropical species are also officially underway. EMBRAPA is leading the complete sequencing of two species of importance for Brazilian aquaculture, within the Animal Genomic Network, tambaqui (*Colossoma macropomum*)

and cachara (*Pseudoplatystoma reticulatum*), along the species already sequenced, such as red-bellied piranha (*Pygocentrus nattereri*) (Schartl et al., 2019) and arapaima (*Arapaima gigas*) (Du et al., 2019).

Tambaqui is a tropical teleost fish native to the Amazon and Orinoco river basins representing over a quarter of the total aquaculture production in Brazil (Ministério da Agricultura, 2016). Previous studies have characterized the LC-PUFA metabolic pathway in tambaqui, showing that all the desaturase and elongase activities required to convert C₁₈ PUFA into LC-PUFA, namely *fads2*, *elovl5*, and *elovl2*, are present (Ferraz et al., 2019). However, the last step of the VLC-PUFA elongation mediated by *Elovl4* has not been established, yet. Here, we aim to characterize the repertoire of *elovl4* genes in tambaqui and investigate the functions of the putative enzymes by heterologous expression in yeast. This study reveals new insights in the understanding of VLC-PUFA metabolism in teleosts. Despite tambaqui being widely cultured in Brazil, its genome resources are still limited. As part of this study, we also provide the first draft genome assembly of *C. macropomum*, which was an essential tool to identify *elovl4b*, and should be useful for further functional and physiology studies.

2. Materials and methods

2.1. Sample collection, DNA and RNA extraction

Juvenile tambaqui (~65 g) were obtained from a commercial farm in Laranja da Terra, Espírito Santo, Brazil (19°55'45.9"S 40°57'42.1"W). Sections of liver and brain were collected from each fish, preserved in RNA stabilization buffer (3.6 M ammonium sulphate, 18 mM sodium citrate, 15 mM EDTA, pH 5.2), and stored at -80 °C prior to DNA and RNA extraction. Total genomic DNA was isolated from ~25 mg of liver using the QIAamp tissue kit (QIAGEN, Netherlands), with DNA resuspended in water. DNA quality was verified in a 2% agarose gel. Total DNA concentration was estimated to be 607 ng/μl, using Quant-it™ RiboGreen® (Invitrogen, California, USA). A BiotTek® microplate reader was used to determine the quality of the DNA. Total RNA was isolated from ~25 mg of brain using the Illustra RNAspin Mini Kit (GE Healthcare, Chicago, USA) following the manufacturer's instructions. On column DNase digestion was used during

the extraction procedure. RNA quality was verified in 1% agarose gels. Total RNA concentration was estimated using the Quant-it™ RiboGreen® RNA Assay Kit (Invitrogen) and BiotTek® microplate reader to determine sample absorbance.

2.2. Draft genome assembly

The DNA sample of liver was sequenced using the Illumina HiSeq X Ten platform (Macrogen, Seoul, Korea). Overall, a total of 90 Gbp of raw data (2x150 paired-end raw reads) was sequenced. An initial data assessment was performed using FastQC (v.0.11.8) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Then, Trimmomatic (v.0.38) (Bolger et al., 2014) was used to filter and remove the low quality reads and adaptors (Parameters ;LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:50). Notwithstanding, Lighter tool (v.1.1.1) (Song et al., 2014) was still applied, with parameters (-k 17 1500000000 0.1), to correct artefact errors from sequencing. The final step of raw dataset validation was done with GenomeScope (v.1.0.0) (Vurture et al., 2017) software (K-mer; 27, 29, 31) to estimate various parameters of *C. macropomum* genome (size, heterozygosity rate, and repeated content) from the clean short reads, allowing to obtain a first overview of the species general proprieties.

The *de novo* draft genome assembly was performed using the W2RAP pipeline (v.0.1) (Clavijo et al., 2017), under the default settings, and following the authors' protocol, available online (<https://www.biorxiv.org/content/10.1101/110999v1>). Briefly, the KAT (v.2.4.1) (Mapleson et al., 2016) software was used to evaluate the k-mer coverage distribution and determine the cut-off frequency suitable to W2RAP pipeline. After that, W2RAP pipeline was run in default mode under the specifications (-t 30 -k 200 -m 400 -min_freq 5 -d 32 --dump_all 1). In the end of the assembly process, the draft genome assembly was assessed using the QUAST software (v.5.0.2) (Gurevich et al., 2013) to provide the general metrics, and the Benchmarking Universal Single-Copy Orthologs (BUSCO (v.3.0.2)) (Simão et al., 2015), using several databases (Eukariota, Metazoa, Vertebrata, and Actinopterygii), to verify the gene content completeness.

2.3. Identification of tambaqui *elovl4a* and *elovl4b* gene sequences

A combination of approaches was used to obtain *elovl4a* and *elovl4b* gene sequences. Initially, the sequences from the *Pygocentrus nattereri elovl4a* (XM_017710341.1) and *elovl4b* (XM_017722369.1) were used for blast searches of liver and brain RNA-seq tambaqui projects. In the case of *elovl4a*, we were able to deduce the full open reading frame (ORF) from the brain RNA-Seq dataset (Ferraz et al., 2020; unpublished). This blast search produced a single hit (TRINITY_DN10862), which was later used to design primers for amplification of the ORF length by PCR. All primers were designed with Primer3 (v. 0.4.0) software (<http://bioinfo.ut.ee/primer3-0.4.0/>) and manufactured by STABVIDA (Portugal). For PCR, cDNA *C. macropomum* brain was synthesized from total RNA (~30 µg) extracted from brain using NZY First-Strand cDNA Synthesis Kit (NzyTech, Portugal) following manufacturer's protocol, posteriorly Flash High-Fidelity PCR Master Mix was used (Thermo Fisher Scientific, Waltham, MA, USA) to amplify cDNA templates. Primer sequences and PCR conditions used in the present study are described in Supplementary Table 1. PCR products were run on 1% agarose gels and those with the expected sizes were excised and purified (Thermo Fisher Scientific) for further sequencing (GATC Biotech, Constance, Germany).

Unlike *elovl4a*, blast searches to RNA-Seq projects were unsuccessful for the tambaqui *elovl4b*, probably due to the lack of relevant gene expression in the examined available tissues (brain and liver). Similarly, we were also unable to amplify the first fragment of *elovl4b* by PCR on brain and liver cDNA as template, and using degenerate primers designed on conserved regions of *elovl4b* from *P. nattereri* (XM_017722369.1), *D. rerio* (NM_199972.1) and *S. salar* (NM_001195552.1). Alternatively, the nucleotide sequence of *elovl4b* was obtained from blast to a genome draft assembly of *C. macropomum*, which is also presented in this work. Briefly, the *P. nattereri elovl4b* gene sequence (XM_017722369.1) was used to blast search a *C. macropomum* draft genome. We identified a scaffold (79611) with the highest sequence similarity to the *P. nattereri elovl4b* gene. A contiguous sequence was manually curated to identify the exonic portions through comparison to *P. nattereri elovl4b* gene. Introns were removed to construct the presumed coding sequence (CDS). The deduced amino acid sequence was compared to the corresponding orthologs from other species. The predicted CDS of the

tambaqui *elovl4b*, identified in the genome scaffold 79611, was synthesized by Integrated DNA Technologies, BVBA (Belgium), as an insert into the pUC19 plasmid. The exon-intron structure for other ElovI genes was determined following the process described above for ElovI4.

2.4. Phylogenetic analysis

To confirm the orthology of the isolated genes, the deduced amino acid sequences of tambaqui ElovI4 were used to construct a phylogenetic tree. For this purpose, sequences of ElovI4a, ElovI4b, ElovI2 and ElovI5 from several Actinopterygii and other vertebrates such as Chondrichthyes and Tetrapoda were aligned with MUSCLE (default settings). This analysis involved 35 amino acid sequences. A total of 372 positions were considered in the final dataset, with complete deletion of positions containing gaps. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-7824.23) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model. Statistical support was calculated using the Bootstrap method in a total of 1000 repetitions. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The evolutionary analyses were performed in MEGA X (Kumar et al., 2018).

2.5. Functional characterization of ElovI4a and ElovI4b

The tambaqui ElovI4a and ElovI4b were functionally characterized in yeast. Briefly, the ORF of the *elovl4a* genes were amplified with primers containing restriction sites *Xba*I (forward) and *Bam*HI (reverse) (Supplementary Table 1), using brain cDNA and then digested. Thus, *elovl4b* was isolated from the direct digestion of the pUC19-*elovl4b* with the enzymes *Hind*III (forward) and *Xba*I (reverse). Subsequently both *elovl4* were purified to obtain ORF that were subsequently ligated into the similarly restricted yeast expression pYES2 (Thermo Fisher Scientific) to produce the constructs pYES2-*elovl4a* and pYES2-*elovl4b*. Finally, all constructs were confirmed by Sanger sequencing (GATC Biotech).

Yeast competent cells InvSc1 (Invitrogen) were transformed with pYES2-*elovl4a* and pYES2-*elovl4b* using the S.c. EasyComp™ Transformation Kit (Invitrogen). Selection of yeast containing the pYES2 constructs was done on *S. cerevisiae* minimal medium minus uracil (SCMM^{-ura}) plates. In order to determine the role of the tambaqui Elov14 in the biosynthesis of VLC-PUFA, a single colony of transgenic yeast expression either *elovl4a* or *elovl4b* was grown at 30 °C in SCMM^{-ura} broth with one of the following PUFA substrates: 18:3n-3, 18:2n-6, 18:4n-3, 18:3n-6, 20:5n-3, 20:4n-6, 22:5n-3, 22:4n-6 and 22:6n-3. In order to compensate for differences in substrate uptake efficiency by yeast, final PUFA substrate concentrations varied according to their fatty acyl chain lengths as 0.5 mM (C₁₈), 0.75 mM (C₂₀) and 1.0 mM (C₂₂) (Lopes-Marques et al., 2017). Yeast were grown in the presence of PUFA substrates for 2 days at 30 °C, and then harvested, washed twice with 5 mL ice-cold HBSS (Invitrogen) containing 1 % (w/v) fatty acid free bovine serum albumin (Sigma-Aldrich), and freeze dried for 24 h for further analyses.

Total lipids extracted from yeast (Folch et al., 1957) were used to prepare fatty acyl methyl esters (FAME) that were further analyzed by gas chromatography as previously described (Li et al., 2017b). FAME were identified based on retention times using an Fisons GC-8160 (Thermo Fisher Scientific) gas chromatograph equipped with a 60 m × 0.32 mm i.d. × 0.25 µm ZB-wax column (Phenomenex, Macclesfield, UK) and flame ionization detector. The elongation conversion efficiencies of exogenously added PUFA substrates were calculated as [areas of first product and longer chain products/(areas of all products with longer chain than substrate + substrate area)] × 100.

In order to test the role of the tambaqui Elov14 in VLC-SFA biosynthesis, yeast transformed with pYES2-*elovl4a*, pYES2-*elovl4b*, or empty pYES2 (control), were incubated in triplicate in the absence of exogenously added substrates. Yeast culture, harvest and collection was carried out as described for yeast supplemented with PUFA substrates. The elongation efficiency towards saturated FA of the tambaqui Elov14 was estimated by comparing the saturated FA profiles of the control yeast (i.e. transformed with empty pYES2) and those of yeast transformed with pYES2-*elovl4a* and pYES2-*elovl4b*.

2.6. Statistical analysis

For the Elov14 functional characterization, the percentages of selected saturated fatty acids of yeast expressing the tambaqui *elov14a* and *elov14b* were compared with those of control yeast transformed with the empty pYES2 vector, by a Student's t-test ($P < 0.05$) (SPSS version 11.5, SPSS, Chicago, IL, USA). FA contents in yeast were presented as mean $\pm$ SD (n=3).

3. Results

3.1. A draft genome assembly of tambaqui

As part of a larger study addressing the nutritional physiology of tambaqui, we generated a novel genomic resource. A draft/low coverage (22x) and cost-effective assembly of tambaqui's genome was produced. About 491 million clean and low error rate pair-end (PE) short reads (*error rate* < 1%) were generated (Table 1). These reads allowed the estimation of the genome size of 2.3 Gb – higher than previous estimates (Hardie and Hebert 2004, 2003; Lobo 2015), a heterozygosity rate between 0.08-0.09 % and between 24-28 % repeat content (Supplementary Table 2). The present draft assembly has a total length of 2.31 Gb, a N50 contig length of 18 kb and a Scaffold N50 length of 25 kb. In addition, and although relatively fragmented, this draft assembly showed a high level of gene content in BUSCO analyses (e.g. % of total genes found (complete + fragmented) in Metazoa and Vertebrata databases were 95.6 % and 96.1 %, respectively) (Table 1). All the data resources generated in this work were submitted to GenBank (details can be checked in Supplementary Table 3).

3.2. Elov14 sequences and phylogenetic analyses

Using various experimental strategies, we were able to deduce the putative full ORF of tambaqui *elov14a* and *elov14b*. In the case of *elov14a*, the inspection of two RNA-Seq projects allowed a clear identification of a single sequence with similarity to this gene. In the case of *elov14b*, various attempts were made but without match hits. In effect, *elov14b* has a far more restricted tissue expression pattern (e.g. eye) and it is probably absent from the analyzed RNA-Seq projects (liver and brain). To explore the possibility that *elov14b* was independently lost in this species, we next investigated a draft genome assembly. By applying various blast strategies, we were able to identify two positive and

different hits (Figure 1A). A careful examination of both scaffolds allowed the identification of the *C. macropomum elovl4b*. Both *elovl4a* and *elovl4b* have of 948 and 912 bp, respectively, encoding putative proteins of 315 and 303 amino acids. Accession numbers for the herein isolated *elovl4a* and *elovl4b* genes orthologs have been deposited to GenBank as NM936179 and NM947651, respectively. Both genes (*elovl4a* and *elovl4b*) are organized in six exons, with the first five having the same number of base pairs in both genes, and the last exon containing 312 and 276 bp in *elovl4a* and *elovl4b*, respectively (Figure 1A). Exon organization of the tambaqui *elovl2* and *elovl5* are also shown in Figure 1A.

A phylogenetic tree was constructed based on the amino acid sequence alignment of the tambaqui Elov4a and Elov4b deduced proteins with other teleost fish Elov4 (Figure 1B). The phylogenetic analysis showed that the tambaqui Elov4a and Elov4b grouped together with their corresponding teleost orthologs. Moreover, sequence analysis shows that both tambaqui Elov4 possess the conserved diagnostic histidine box HXXHH motif present in all elongases and the endoplasmic reticulum (ER) retrieval signal RXKXX (Elov4a) and KXKXX (Elov4b) at the carboxyl terminus, which could be observed in the alignment made with *D. rerio* and *P. nattereri* (Figure 2). Regarding the diagnostic “Q” (glutamine) in position –5 from the H**HH, characteristic of PUFA elongases, this is only present in Elov4b sequence. Elov4a does not display this feature, similarly to *P. nattereri* Elov4 (Figure 2).

3.3. Roles of the tambaqui Elov4a and Elov4b in the biosynthesis of VLC-PUFA and VLC-SFA

The activity of the tambaqui Elov4 enzymes was investigated by growing yeast transformed with pYES2-*elovl4a* and pYES2-*elovl4b* in the presence of potential PUFA with acyl chains of C₁₈ (18:3n-3, 18:2n-6, 18:4n-3 and 18:3n-6), C₂₀ (20:5n-3 and 20:4n-6) and C₂₂ (22:5n-3, 22:4n-6 and 22:6n-3). The results showed that both tambaqui Elov4 enzymes can contribute to VLC-PUFA biosynthesis since yeast expressing both *elovl4a* and *elovl4b* were able to elongate a large variety of PUFA substrates (Table 2). The step-wise elongation reactions lead to the production of a series of polyenes that, in the case of Elov4a, reached up to C₃₆ long for all PUFA substrates tested except for 22:4n-6 and DHA (Table 2). Polyenes of up to C₃₄ were detected in yeast expression the tambaqui

elovl4b (Table 2). It is noteworthy that Elov4a and Elov4b were able to elongate both 20:5n-3 and 22:5n-3 up to 24:5n-3, the substrate for DHA biosynthesis via the Sprecher pathway (Table 2).

The putative Elov4 enzymes of tambaqui were functionally characterized by comparing the FA profiles of control *S. cerevisiae* (i.e. transformed with empty pYES2) with those of yeast expressing either tambaqui *elovl4a* or *elovl4b*. Results showed that pYES2-*elovl4a* and pYES2-*elovl4b* transformed yeast contained higher proportions of saturated FA from 20:0 to 32:0 than their corresponding controls (Table 3). However, only the results for *elovl4b* with 28:0 and 30:0 as substrates, were significantly higher in comparison to control yeast (Table 3). These results confirmed that Elov4 is involved in the biosynthesis of VLC-SFA.

4. Discussion

The importance of LC-PUFA is already known for homeostasis and better fish development (Tocher, 2010). However, recently, several studies on commercially important fish species have emphasized the importance of very long-chain saturated fatty acids (VLC-SFA) and PUFA (VLC-PUFA) in aquaculture (Carmona-Antonanzas et al., 2011; Jin et al., 2017; Oboh et al., 2017a; Torres et al., 2020). Recently, the LC-PUFA biosynthesis pathway in tambaqui was unveiled (Ferraz et al., 2019) showing that all the desaturase and elongase activities required to convert C₁₈ PUFA into LC-PUFA are present. However, the repertoire and function of the tambaqui *elovl4* had not been investigated.

The knowledge of genomics has evolved remarkably in recent years. The rapid development of sequencing technologies has allowed the characterization of a large number of genomes in species of interest using next-generation technologies. Here, as part of a larger study addressing the nutritional physiology of tambaqui, a novel genomic resource was generated. The key enzymes of VLC-FA biosynthesis were investigated in this draft genome. Due to the high completeness and the long scaffold N50 of the tambaqui draft genome assembly, it was possible to identify the *elovl4b* gene where

other methodological approaches (e.g. PCR with degenerate primers) had failed. Such a goal had not been possible to achieve by searches within the transcriptomes available from liver and brain. Furthermore, these results helped us to clarify that *elovl4b* has not been independently lost in tambaqui.

The repertoire and function of genes encoding Elovl4 proteins was investigated, and from then on, the phylogenetic analysis of Elovl4 recognized that *C. macropomum* has two *elovl4*-like elongases that have high homology to previously characterized Elovl4 from teleosts. They were termed accordingly Elovl4a and Elovl4b. This finding is in agreement with *in silico* studies suggesting that virtually all teleosts possess both types of *elovl4* genes (Castro et al., 2016). Both *C. macropomum elovl4a* and *elovl4b* encoded putative proteins that contained a diagnostic histidine box, indicating that Elovl enzymes have functional regions highly conserved during evolution (Hashimoto et al., 2008). Notably, the Elovl4b sequence contains the diagnostic “Q” (glutamine) in position –5 from the H**HH motif, characteristic of PUFA elongases (Hashimoto et al., 2008). Interestingly, although Elovl4a does not contain the diagnostic “Q”, our functional characterization data demonstrated that this enzyme can operate toward PUFA substrates, and so participates in the biosynthesis of VLC-PUFA.

Both *C. macropomum* Elovl4 elongases characterized in the present study were demonstrated to elongate PUFA substrates including compounds of different chain lengths (C_{18–22}) and series (n-3 and n-6). These results suggest that both Elovl4a and Elovl4b from tambaqui participate in the biosynthesis of VLC-PUFA of up to 34-36 carbons, through consecutive elongations from a varied range of shorter-chain PUFA precursors. These results are in agreement with those reported in *C. gariepinus* and *A. schlegelii* species, in which both their Elovl4 proteins (Elovl4a and Elovl4b) contribute to VLC-PUFA biosynthesis (Jin et al., 2017; Oboh et al., 2017a). In contrast, the pioneer study reporting for first time on functions of Elovl4 from teleosts showed that in *D. rerio* only Elovl4b, but not Elovl4a, plays important roles in the biosynthesis of VLC-PUFA (Monroig et al., 2010). Such conclusion was drawn upon functional analysis data derived from yeast assays similar to those used herein, along with gene expression analyses, that demonstrated that the *D. rerio elovl4b* is highly active in retina and pineal gland

(Monroig et al., 2010), tissues containing photoreceptors in which VLC-PUFA accumulate (Agbaga et al., 2010; Deak et al., 2019).

Vertebrate Elovl4 have demonstrated roles in the biosynthesis of VLC-SFA (Deak et al., 2019). Here, yeast expressing both *elovl4a* and *elovl4b* had increased levels of VLC-SFA of C_{26–30} as compared to control yeast, in agreement with elongation abilities of some teleost Elovl4 reported previously (Carmona-Antonanzas et al., 2011; Li et al., 2017a; Li et al., 2017b; Monroig et al., 2010; Monroig et al., 2012; Monroig et al., 2011; Oboh et al., 2017a; Ran et al., 2019). However, such increased production of VLC-SFA was only statistically significant for Elovl4b, more specifically for 28:0 and 30:0, but not for Elovl4a. These results differed from those reported in *D. rerio* since its Elovl4a but not Elovl4b, was found to play major roles in VLC-SFA (Monroig et al., 2010).

Both VLC-PUFA and VLC-SFA are recognized to be biosynthesized by Elovl4 in a tissue-specific manner (Deak et al., 2019). Provided that *elovl4a* and *elovl4b* are strongly expressed in eye and brain from teleosts (Monroig et al., 2010; Oboh et al., 2017a; Torres et al., 2020), it is reasonable to speculate that Elovl4 are responsible for endogenous supply of these compounds in neural tissues. This hypothesis is partly supported by the identification of certain VLC-PUFA including 24:6n-3, 26:6n-3, 28:6n-3, 30:6n-3, 32:6n-3 and 34:6n-3, in phosphatidylcholine (PC) from teleost retina (Garlito et al., 2019). For instance, Atlantic salmon retinal PC contained detectable levels of 24:6n-3 and 32:6n-3 (Garlito et al., 2019), in agreement with the salmon Elovl4b being able to elongate DHA up to 34:6n-3 (Carmona-Antonanzas et al., 2011). Our functional characterization data, however, indicated that both tambaqui Elovl4 isoforms could only elongate DHA to 24:6n-3, with longer elongation products not being detected. Unfortunately, retinal samples from tambaqui were not available in the present study and hence we were unable to confirm whether, along 24:6n-3, other VLC-PUFA can be endogenously synthesized by tambaqui. More clearly though, our data strongly suggest that tambaqui Elovl4 can actively contribute to the biosynthesis of DHA, a compound that is highly abundant in neural tissues as described above for VLC-PUFA (Bell and Tocher, 1989; Dyll, 2015).

DHA is quantitatively the most important n-3 LC-PUFA in the brain, and has consistently been shown to have unique and indispensable roles in the neuronal

function, being an important constituent of brain phospholipids and playing a role in maintaining structural and functional integrity of membranes in brain (Bell and Tocher, 1989; Dyll, 2015). In addition to dietary origin, DHA can be synthesized endogenously. Arguably, the most common route for DHA biosynthesis is the Sprecher pathway, which involves elongation of 22:5n-3 to 24:5n-3, before consecutive reactions of $\Delta 6$ desaturation and a peroxisome β -oxidation occur (Castro et al., 2016; Monroig and Kabeya, 2018; Monroig et al., 2016). Our results demonstrate that the Elov4 characterized herein from tambaqui can contribute to DHA biosynthesis through this pathway as both isoforms were able to elongate 22:5n-5 to 24:5n-3. Such characteristic, widespread among teleost Elov4 enzymes e.g. African catfish (*C. gariepinus*) (Obboh et al., 2017a) and rainbow trout (*O. mykiss*) (Zhao et al., 2019), had been postulated to compensate for lack of Elov2 in genomes of many teleosts (Castro et al., 2016). Recently, a study on the Atlantic bluefin tuna (*T. thynnus*) Elov4b has shown that this enzyme is indeed able to produce 24:5n-3 before this metabolite can be further desaturated by the Fads2 existing in this species (Betancor et al., 2020). Tambaqui $\Delta 6$ Fads2 was also demonstrated to desaturate 24:5n-3 to 24:6n-3 suggesting that the this species is capable to operate the Sprecher pathway (Ferraz et al., 2019). Importantly, the present study shows that, along Elov2 (Ferraz et al., 2019), key elongation of 22:5n-3 to 24:5n-3 can be also mediated by Elov4 enzymes as evidenced in our in vitro assays. These results indicate that Elov2 (Ferraz et al., 2019) and both Elov4 enzymes characterized in the present study have partly overlapping functions, as previously described in Elov2 and Elov5 in tambaqui itself (Ferraz et al., 2019) or other teleosts (Castro et al., 2016).

In conclusion, a draft genome assembly of tambaqui has been presented. This data collection can provide insight into the nutritional physiology and metabolism of essential FA including VLC-PUFA and VLC-SFA. Moreover, the tambaqui draft genome assembly was useful to identify the *elov4b* gene. The present study confirmed that both *C. macropomum* Elov4 enzymes play critical roles in the biosynthesis of VLC-PUFA and VLC-SFA. Elov4a showed the ability to biosynthesize VLC-PUFA of up to 36 carbons, whereas Elov4b enabled the synthesis of polyenes of up to 34 carbons. Moreover, both

Elovl4 from tambaqui can contribute to the biosynthesis of DHA via the Sprecher pathway since they produced the metabolic intermediate 24:5n-3 from 22:5n-3.

Acknowledgments

This study was supported by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil, COMPETE 2020, Portugal 2020 and the European Union through the ERDF, grant number 031342, and by FCT through national funds.

References

- Agbaga, M.-P., Mandal, M.N.A., Anderson, R.E., 2010. Retinal very long-chain PUFAs: new insights from studies on ELOVL4 protein. *J. Lipid Res.* 51, 1624-1642.
- Bell, M., Tocher, D.R., 1989. Molecular species composition of the major phospholipids in brain and retina from rainbow trout (*Salmo gairdneri*). Occurrence of high levels of di-(n-3) polyunsaturated fatty acid species. *Biochem. J.* 264, 909-915.
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5, 3657.
- Betancor, M.B., Oboh, A., Ortega, A., Mourente, G., Navarro, J.C., de la Gandara, F., Tocher, D.R., Monroig, O., 2020. Molecular and functional characterisation of a putative *elovl4* gene and its expression in response to dietary fatty acid profile in Atlantic bluefin tuna (*Thunnus thynnus*). *Comp. Biochem. Physiol.* 240B, 110372.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Buzzi, M., Henderson, R., Sargent, J., 1996. The desaturation and elongation of linolenic acid and eicosapentaenoic acid by hepatocytes and liver microsomes from rainbow trout (*Oncorhynchus mykiss*) fed diets containing fish oil or olive oil. *Biochim. Biophys. Acta* 1299, 235-244.
- Buzzi, M., Henderson, R.J., Sargent, J.R., 1997. Biosynthesis of docosahexaenoic acid in trout hepatocytes proceeds via 24-carbon intermediates. *Comp. Biochem. Physiol.* 116B, 263-267.
- Calder, P.C., 2014. Very long chain omega-3 (n-3) fatty acids and human health. *Eur. J. Lipid Sci. Technol.* 116, 1280-1300.
- Campoy, C., Escolano-Margarit, M.V., Anjos, T., Szajewska, H., Uauy, R., 2012. Omega 3 fatty acids on child growth, visual acuity and neurodevelopment. *Br. J. Nutr.* 107, S85-S106.

489 Carmona-Antonanzas, G., Monroig, O., Dick, J.R., Davie, A., Tocher, D.R., 2011.
 490 Biosynthesis of very long-chain fatty acids (C>24) in *Atlantic salmon*: cloning,
 491 functional characterisation, and tissue distribution of an Elovl4 elongase. *Comp.*
 492 *Biochem. Physiol.* 159B, 122-129.

493 Castro, L.F., Tocher, D.R., Monroig, O., 2016. Long-chain polyunsaturated fatty acid
 494 biosynthesis in chordates: Insights into the evolution of Fads and Elovl gene
 495 repertoire. *Prog. Lipid Res.* 62, 25-40.

496 Clavijo, B., Accinelli, G.G., Wright, J., Heavens, D., Barr, K., Yanes, L., Di Palma, F., 2017.
 497 W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from
 498 short read data. *BioRxiv* 110999.

499 Conte, M.A., Gammerdinger, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D., 2017. A high
 500 quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the
 501 structure of two sex determination regions. *BMC Genomics* 18, 341.

502 Deak, F., Anderson, R.E., Fessler, J.L., Sherry, D.M., 2019. Novel Cellular Functions of Very
 503 Long Chain-Fatty Acids: Insight From ELOVL4 Mutations. *Front. Cell. Neurosci.* 13,
 504 428.

505 Delgado-Lista, J., Perez-Martinez, P., Lopez-Miranda, J., Perez-Jimenez, F., 2012. Long
 506 chain omega-3 fatty acids and cardiovascular disease: a systematic review. *Br. J. Nutr.*
 507 107, S201-S213.

508 Du, K., Wuertz, S., Adolphi, M., Kneitz, S., Stöck, M., Oliveira, M., Nóbrega, R., Ormanns,
 509 J., Kloas, W., Feron, R., 2019. The genome of the arapaima (*Arapaima gigas*) provides
 510 insights into gigantism, fast growth and chromosomal sex determination system. *Sci.*
 511 *Rep.* 9, 5293.

512 Dyall, S.C., 2015. Long-chain omega-3 fatty acids and the brain: a review of the
 513 independent and shared effects of EPA, DPA and DHA. *Front. Aging Neurosci.* 7, 52.

514 Ferraz, R.B., Kabeya, N., Lopes-Marques, M., Machado, A.M., Ribeiro, R.A., Salaro, A.L.,
 515 Ozório, R., Castro, L.F.C., Monroig, Ó., 2019. A complete enzymatic capacity for long-
 516 chain polyunsaturated fatty acid biosynthesis is present in the Amazonian teleost
 517 tambaqui, *Colossoma macropomum*. *Comp. Biochem. Physiol.* 227B, 90-97.

518 Folch, J., Lees, M., Sloane Stanley, G., 1957. A simple method for the isolation and
519 purification of total lipids from animal tissues. *J. Biol. Chem.* 226, 497-509.

520 Garlito, B., Portolés, T., Niessen, W., Navarro, J., Hontoria, F., Monroig, Ó., Varó, I.,
521 Serrano, R., 2019. Identification of very long-chain (>C24) fatty acid methyl esters
522 using gas chromatography coupled to quadrupole/time-of-flight mass spectrometry
523 with atmospheric pressure chemical ionization source. *Anal. Chim. Acta.* 1051, 103-
524 109.

525 Guillou, H., Zadravec, D., Martin, P.G., Jacobsson, A., 2010. The key roles of elongases
526 and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice.
527 *Prog. Lipid Res.* 49, 186-199.

528 Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUASt: quality assessment tool
529 for genome assemblies. *Bioinformatics* 29, 1072-1075.

530 Hashimoto, K., Yoshizawa, A.C., Okuda, S., Kuma, K., Goto, S., Kanehisa, M., 2008. The
531 repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic
532 genomes. *J. Lipid Res.* 49, 183-191.

533 Innis, S.M., 2007. Dietary (n-3) fatty acids and brain development. *J. Nutr.* 137, 855-859.

534 Ishikawa, A., Kabeya, N., Ikeya, K., Kakioka, R., Cech, J.N., Osada, N., Leal, M.C., Inoue, J.,
535 Kume, M., Toyoda, A., Tezuka, A., Nagano, A.J., Yamasaki, Y.Y., Suzuki, Y., Kokita, T.,
536 Takahashi, H., Lucek, K., Marques, D., Takehana, Y., Naruse, K., Mori, S., Monroig, Ó.,
537 Ladd, N., Schubert, C.J., Matthews, B., Peichel, C.L., Seehausen, O., Yoshizaki, G.,
538 Kitano, J., 2019. A key metabolic gene for recurrent freshwater colonization and
539 radiation in fishes. *Science* 364, 886-889.

540 Jakobsson, A., Westerberg, R., Jacobsson, A., 2006. Fatty acid elongases in mammals:
541 their regulation and roles in metabolism. *Prog. Lipid Res.* 45, 237-249.

542 Jin, M., Monroig, O., Navarro, J.C., Tocher, D.R., Zhou, Q.-C., 2017. Molecular and
543 functional characterisation of two elovl4 elongases involved in the biosynthesis of
544 very long-chain (> C24) polyunsaturated fatty acids in black seabream *Acanthopagrus*
545 *schlegelii*. *Comp. Biochem. Physiol.* 212B, 41-50.

546 Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data
547 matrices from protein sequences. *Bioinformatics* 8, 275-282.

548 Kabeya, N., Yamamoto, Y., Cummins, S.F., Elizur, A., Yazawa, R., Takeuchi, Y., Haga, Y.,
549 Satoh, S., Yoshizaki, G., 2015. Polyunsaturated fatty acid metabolism in a marine
550 teleost, Nibe croaker *Nibea mitsukurii*: Functional characterization of Fads2
551 desaturase and Elovl5 and Elovl4 elongases. *Comp. Biochem. Physiol.* 188B, 37-45.

552 Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular
553 evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547-
554 1549.

555 Li, S., Monroig, Ó., Navarro, J.C., Yuan, Y., Xu, W., Mai, K., Tocher, D.R., Ai, Q., 2017a.
556 Molecular cloning and functional characterization of a putative Elovl4 gene and its
557 expression in response to dietary fatty acid profiles in *orange-spotted grouper* *E*
558 *pinephelus coioides*. *Aquaculture Res.* 48, 537-552.

559 Li, S., Monroig, Ó., Wang, T., Yuan, Y., Navarro, J.C., Hontoria, F., Liao, K., Tocher, D.R.,
560 Mai, K., Xu, W., 2017b. Functional characterization and differential nutritional
561 regulation of putative Elovl5 and Elovl4 elongases in large yellow croaker
562 (*Larimichthys crocea*). *Sci. Rep.* 7, 2303.

563 Li, Y., Wen, Z., You, C., Xie, Z., Tocher, D.R., Zhang, Y., Wang, S., Li, Y., 2020. Genome
564 wide identification and functional characterization of two LC-PUFA biosynthesis
565 elongase (*elovl8*) genes in rabbitfish (*Siganus canaliculatus*). *Aquaculture* 735127.

566 Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong,
567 J.S., Minkley, D.R., Zimin, A., 2016. The Atlantic salmon genome provides insights into
568 rediploidization. *Nature* 533, 200.

569 Lopes-Marques, M., Ozorio, R., Amaral, R., Tocher, D.R., Monroig, O., Castro, L.F., 2017.
570 Molecular and functional characterization of a *fads2* orthologue in the Amazonian
571 teleost, *Arapaima gigas*. *Comp. Biochem. Physiol.* 203B, 84-91.

572 Louro, B., De Moro, G., Garcia, C., Cox, C.J., Veríssimo, A., Sabatino, S.J., Santos, A.M.,
573 Canário, A.V., 2019. A haplotype-resolved draft genome of the European sardine
574 (*Sardina pilchardus*). *GigaScience* 8, giz059.

575 Machado, A.M., Tørresen, O.K., Kabeya, N., Couto, A., Petersen, B., Felício, M., Campos,
576 P.F., Fonseca, E., Bandarra, N., Lopes-Marques, M., 2018. “Out of the Can”: A Draft
577 Genome Assembly, Liver Transcriptome and Nutrigenomics of the European Sardine,
578 *Sardina pilchardus*. Genes 9.10, 485.

579 Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., Clavijo, B.J., 2016. KAT:
580 a K-mer analysis toolkit to quality control NGS datasets and genome assemblies.
581 Bioinformatics 33, 574-576.

582 Ministério da Agricultura, Brasil., 2016. Produção da pecuária municipal / IBGE. Accessed
583 2017/10. [http://biblioteca.ibge.gov.br/index.php/biblioteca-](http://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=784)
584 [catalogo?view=detalhes&id=784](http://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=784).

585 Monroig, Ó., Kabeya, N., 2018. Desaturases and elongases involved in polyunsaturated
586 fatty acid biosynthesis in aquatic invertebrates: a comprehensive review. Fish. Sci. 1-
587 18.

588 Monroig, O., Lopes-Marques, M., Navarro, J.C., Hontoria, F., Ruivo, R., Santos, M.M.,
589 Venkatesh, B., Tocher, D.R., Castro, L.F., 2016. Evolutionary functional elaboration of
590 the Elovl2/5 gene family in chordates. Sci. Rep. 6, 20510.

591 Monroig, O., Rotllant, J., Cerda-Reverter, J.M., Dick, J.R., Figueras, A., Tocher, D.R., 2010.
592 Expression and role of Elovl4 elongases in biosynthesis of very long-chain fatty acids
593 during zebrafish *Danio rerio* early embryonic development. Biochim. Biophys. Acta.
594 1801, 1145-1154.

595 Monroig, Ó., Tocher, D.R., Hontoria, F., Navarro, J.C., 2013. Functional characterisation
596 of a Fads2 fatty acyl desaturase with $\Delta 6/\Delta 8$ activity and an Elovl5 with C16, C18 and
597 C20 elongase activity in the anadromous teleost meagre (*Argyrosomus regius*).
598 Aquaculture 412-413, 14-22.

599 Monroig, Ó., Wang, S., Zhang, L., You, C., Tocher, D.R., Li, Y., 2012. Elongation of long-
600 chain fatty acids in rabbitfish *Siganus canaliculatus*: Cloning, functional
601 characterisation and tissue distribution of Elovl5- and Elovl4-like elongases.
602 Aquaculture 350-353, 63-70.

- Monroig, Ó., Webb, K., Ibarra-Castro, L., Holt, G.J., Tocher, D.R., 2011. Biosynthesis of long-chain polyunsaturated fatty acids in marine fish: Characterization of an Elov14-like elongase from cobia *Rachycentron canadum* and activation of the pathway during early life stages. *Aquaculture* 312, 145-153.
- Oboh, A., Kabeya, N., Carmona-Antoñanzas, G., Castro, L.F.C., Dick, J.R., Tocher, D.R., Monroig, O., 2017b. Two alternative pathways for docosahexaenoic acid (DHA, 22:6n-3) biosynthesis are widespread among teleost fish. *Sci. Rep.* 7, 3889.
- Oboh, A., Navarro, J.C., Tocher, D.R., Monroig, O., 2017a. Elongation of very long-chain (> C24) fatty acids in *Clarias gariepinus*: Cloning, functional characterization and tissue expression of *elov14* elongases. *Lipids* 52, 837-848.
- Palti, Y., Genet, C., Gao, G., Hu, Y., You, F.M., Boussaha, M., Rexroad, C.E., Luo, M.-C., 2012. A second generation integrated map of the rainbow trout (*Oncorhynchus mykiss*) genome: analysis of conserved synteny with model fish genomes. *Mar. Biotechnol.* 14, 343-357.
- Ran, Z., Xu, J., Liao, K., Monroig, Ó., Navarro, J.C., Oboh, A., Jin, M., Zhou, Q., Zhou, C., Tocher, D.R., 2019. Biosynthesis of long-chain polyunsaturated fatty acids in the razor clam *Sinonovacula constricta*: Characterization of four fatty acyl elongases and a novel desaturase capacity. *Biochim. Biophys. Acta* 1864, 1083-1090.
- Schartl, M., Kneitz, S., Volkoff, H., Adolphi, M., Schmidt, C., Fischer, P., Minx, P., Tomlinson, C., Meyer, A., Warren, W.C., 2019. The Piranha Genome Provides Molecular Insight Associated to Its Unique Feeding Behavior. *Genome biology and evolution. Genome Biol. Evol.* 11, 2099-2106.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210-3212.
- Song, L., Florea, L., Langmead, B., 2014. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Boil.* 15, 509.
- Sprecher, H., 2000. Metabolism of highly unsaturated n-3 and n-6 fatty acids. *Biochim. Biophys. Acta* 1486, 219-231.

- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S., Hecht, J., Knaust, F., Belkhir, K., Klages, S., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* 5, 5770.
- Tocher, D.R., 2010. Fatty acid requirements in ontogeny of marine and freshwater fish. *Aquaculture Res.* 41, 717-732.
- Tocher, D.R., 2015. Omega-3 long-chain polyunsaturated fatty acids and aquaculture in perspective. *Aquaculture* 449, 94-107.
- Tocher, D.R., Agaba, M.K., Hastings, N., Teale, A.J., 2003. Biochemical and molecular studies of the polyunsaturated fatty acid desaturation pathway in fish. In: Browman H I, Skiftesvik A B (ed.). *The Big Fish Bang: Proceedings of the 26th Annual Larval Fish Conference*, Bergen, Norway: Institute of Marine Research (IMR) / Fishlarvae.com, pp. 211-228.
- Torres, M., Navarro, J., Varó, I., Agulleiro, M., Morais, S., Monroig, Ó., Hontoria, F., 2020. Expression of genes related to long-chain (C18–22) and very long-chain (> C24) fatty acid biosynthesis in gilthead seabream (*Sparus aurata*) and Senegalese sole (*Solea senegalensis*) larvae: Investigating early ontogeny and nutritional regulation. *Aquaculture* 734949.
- Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202-2204.
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., 2014. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genetics* 46, 1212.
- Yan, J., Liang, X., Cui, Y., Cao, X., Gao, J., 2017. Elovl4 can effectively elongate C18 polyunsaturated fatty acids in loach *Misgurnus anguillicaudatus*. *Biochem. Biophys. Res. Commun.* 495.4, 2637-2642.
- Zhao, N., Monroig, Ó., Navarro, J.C., Xiang, X., Li, Y., Du, J., Li, J., Xu, W., Mai, K., Ai, Q., 2019. Molecular cloning, functional characterization and nutritional regulation of two

661 *elovl4b* elongases from rainbow trout (*Oncorhynchus mykiss*). Aquaculture 511,
662 734221.

663

668 **Figure 1: A.** Genomic loci and exon organization of the *C. macropomum elovl* genes
 669 involved in PUFA metabolism (*elovl4a*, *elovl4b*, *elovl2* and *elovl5*). **B.** Maximum
 670 likelihood phylogenetic tree of *C. macropomum* Elov14a, Elov14b, Elov12 and Elov15 with
 671 elongase proteins from other vertebrate species. Numbers at nodes indicate branch
 672 support calculated using Bootstrap method with 1000 replications. *C. macropomum*
 673 sequences (Elov14a and Elov14b) studied herein are highlighted. Accession numbers for
 674 all Elov1 sequences are shown in front of each corresponding species.

Figure 2. Comparison of the deduced *C. macropomum* amino acid (AA) sequences of Elovl4a and Elovl4b with those of other species. **A** - For Elovl4a *D. rerio* (NP_957090.1), *P. nattereri* (XP_017565830.1), *O. niloticus* (XP_003443720.1) and *C. macropomum* (this work) were used. **B** - For Elovl4b *D. rerio* (NP_956266.1), *P. nattereri* (XP_017574113.1), *O. niloticus* (XP_003440669.1) and *C. macropomum* (this work). Both AA sequences were aligned using BioEdit. Identical residues are shaded black and similar residues are shaded grey. Indicated are the conserved HXXHH histidine box motif and the ER retrieval signal, where RXKXX (Elovl4a) and KXKXX (Elovl4b) at the carboxyl terminus; “*” represents “Q” (glutamine) in position –5 from the H**HH.

TABLES

Table 1: General stats of *Colossoma macropomum* draft genome assembly.

	Contig *	Scaffold *
Raw Data Stats		
Raw sequencing reads (PE)		491416754
Clean reads		526280086
Assembly Stats		
Total number of Sequences	255110	229835
Total number of Sequences (≥ 1000)	228808	203533
Total number of Sequences (≥ 10000)	71024	62135
Total number of Sequences (≥ 25000)	20829	24453
Total number of Sequences (≥ 50000)	4287	7519
Total length, bp	2311578259	2314112053
Total length (≥ 1000), bp	2292711909	2295245703
Total length (≥ 10000), bp	1638446955	1760715179
Total length (≥ 25000), bp	856096545	1158738469
Total length (≥ 50000), bp	294364635	572298158
N50 length, bp	18011	25040
L50 length, bp	35010	24386
Maximum length, bp	228241	330952
GC content, %	39,36	39,36
Genome coverage, ×	-	22x
Back Mapping Reads, %	-	98%
BUSCOS		
Buscos – Eukariota**	-	C:76.9%[S:31.7%,D:45.2%],F:11.2%,M:11.9%,n:303
Buscos – Metazoa**	-	C:86.7%[S:35.3%,D:51.4%],F:8.9%,M:4.4%,n:978
Buscos – Vetebrata**	-	C:80.1%[S:34.1%,D:46.0%],F:16.0%,M:3.9%,n:2586
Buscos – Actinopterygii**	-	C:77.9%[S:34.3%,D:43.6%],F:11.5%,M:10.6%,n:4584

* All statistics are based on contigs/scaffolds of size ≥500bp.

** (C : Complete Buscos [S: Complete and single copy, %;D:Complete and duplicated, %], F: Fragmented, %, M: Missing, %, n: Number of sequences in Database).

Table 2: Functional characterisation of Elovl4b elongase of *Colossoma macropomum* by heterologous expression in the yeast *Saccharomyces cerevisiae*. Data are presented as the percentage conversions of polyunsaturated fatty acid (FA) substrates calculated according to the formula [areas of first product and longer chain products / (areas of all products with longer chain than substrate + substrate area)] × 100.

FA substrate	FA product	% Conversion	
		Elovl4a	Elovl4b
18:3n-3	20:3n-3	10.9	16.6
	22:3n-3	10.4	9.7
	24:3n-3	35.9	16.9
	26:3n-3	86.1	87.3
	28:3n-3	95.7	100.0
	30:3n-3	98.8	93.6
	32:3n-3	97.0	61.9
	34:3n-3	85.3	N.D.
	36:3n-3	21.2	N.D.
18:2n-6	20:2n-6	6.7	12.1
	22:2n-6	19.9	17.4
	24:2n-6	30.7	25.1
	26:2n-6	83.9	86.1
	28:2n-6	94.3	90.6
	30:2n-6	97.9	86.9
	32:2n-6	90.7	9.5
	34:2n-6	54.0	N.D.
	36:2n-6	12.4	N.D.
18:4n-3	20:4n-3	5.7	8.1
	22:4n-3	15.4	14.4
	24:4n-3	43.5	10.2
	26:4n-3	70.1	100.0
	28:4n-3	94.3	100.0
	30:4n-3	98.7	76.8
	32:4n-3	93.3	77.8
	34:4n-3	87.2	N.D.
	36:4n-3	17.5	N.D.
18:3n-6	20:3n-6	6.7	13.0
	22:3n-6	29.8	26.9
	24:3n-6	57.9	22.2
	26:3n-6	75.9	85.2
	28:3n-6	95.3	93.7
	30:3n-6	99.4	93.5
	32:3n-6	97.4	25.6
	34:3n-6	74.3	N.D.

	36:3n-6	11.2	N.D.
20:5n-3	22:5n-3	17.2	23.8
	24:5n-3	27.1	26.2
	26:5n-3	31.4	19.6
	28:5n-3	71.7	100.0
	30:5n-3	99.0	91.6
	32:5n-3	97.9	88.7
	34:5n-3	92.6	24.5
	36:5n-3	44.9	N.D.
20:4n-6	22:4n-6	16.1	30.8
	24:4n-6	31.4	37.9
	26:4n-6	44.5	55.7
	28:4n-6	75.0	91.6
	30:4n-6	99.0	97.5
	32:4n-6	96.2	67.9
	34:4n-6	88.2	4.8
	36:4n-6	29.4	N.D.
22:5n-3	24:5n-3	20.0	10.0
	26:5n-3	33.4	44.1
	28:5n-3	70.6	94.9
	30:5n-3	98.9	99.5
	32:5n-3	97.0	92.4
	34:5n-3	91.6	19.7
	36:5n-3	40.6	N.D.
22:4n-6	24:4n-6	1.9	6.4
	26:4n-6	N.D.	49.2
	28:4n-6	N.D.	93.0
	30:4n-6	N.D.	98.2
	32:4n-6	N.D.	74.0
	34:4n-6	N.D.	6.7
22:6n-3	24:6n-3	0.1	0.2

700 N.D., not detected

701

Table 3: Selected saturated fatty acids (percentage of total fatty acids) of yeast *Saccharomyces cerevisiae* transformed with either the empty pYES2 vector (Control) or the *Colossoma macropomum elovl4* ORF sequences (pYES2-*elovl4a* or pYES2-*elovl4b*). Results are means $\pm$ SD (n=3). Significant statistical differences between transgenic yeast expressing each *elovl4* gene and its corresponding control are indicated with an asterisk (Student t-test, $P < 0.05$).

SFA	Control	pYES2- <i>elovl4a</i>	P value	Control	pYES2- <i>elovl4b</i>	P value
16:0	61.15 $\pm$ 6.05	49.21 $\pm$ 9.25	0.30	55.41 $\pm$ 3.46	39.68 $\pm$ 2.33	0.37
18:0	35.77 $\pm$ 4.12	39.35 $\pm$ 10.71	0.26	42.22 $\pm$ 3.13	51.32 $\pm$ 3.13	1.00
20:0	0.46 $\pm$ 0.02	0.75 $\pm$ 0.03	0.46	0.39 $\pm$ 0.03	0.87 $\pm$ 0.09	0.31
22:0	0.57 $\pm$ 0.14	1.13 $\pm$ 0.13	0.10	0.60 $\pm$ 0.07	1.03 $\pm$ 0.16	0.12
24:0	0.20 $\pm$ 0.01	0.52 $\pm$ 0.01	0.13	0.26 $\pm$ 0.05	0.53 $\pm$ 0.02	0.31
26:0	1.63 $\pm$ 0.12	5.15 $\pm$ 0.22	0.22	0.98 $\pm$ 0.25	4.36 $\pm$ 0.96	0.06
28:0	0.12 $\pm$ 0.01	2.51 $\pm$ 0.24	0.06	0.09 $\pm$ 0.00	2.00 $\pm$ 0.40	0.04*
30:0	0.05 $\pm$ 0.01	1.13 $\pm$ 0.16	0.20	0.01 $\pm$ 0.00	0.15 $\pm$ 0.01	0.05*
32:0	0.01 $\pm$ 0.00	0.19 $\pm$ 0.02	0.16	0.01 $\pm$ 0.00	0.01 $\pm$ 0.00	0.24

SUPPLEMENTARY MATERIALS

Supplementary Table 1. The primers and the conditions for cloning of *elovl4a* by PCR.

	Primer set function	Primer name	Primer sequence	Cycles	TM	Extension (size bp)
Elov4a	Full ORF	Cma_ELOVL4A_ORF_F	CCCAGAACGGACAATTAGGA	35	60 °C	72 °C/24 s
		Cma_ELOVL4A_ORF_R	GAAGCCTGCCTGTTTCTCTG			
	Restriction site	Cma_ELOVL4A_Pyes_XBAL_F	CCCGGATCCAAGATGGATATTGTAACACATCTCATC	35	62 °C	72 °C/21 s
		Cma_ELOVL4A_Pyes_BAMHI_R	CCCTCTAGACTAGTCACGCTTGGCCCTG			

Supplementary Table 2. GenomeScope (k-mer 27, 29 and 31), Kmergenie and Sga Preqc statistics of *Colossoma macropomum* raw reads.

GenomeScope						
Property	K = 27		K = 29		k = 31	
	min	max	min	max	min	max
Heterozygosity, %	0.0851359	0.0891992	0.0818162	0.0854054	0.0786521	0.0818496
Genome Haploid Length, bp	2,284,669,586	2,286,668,997	2,304,304,586	2,306,199,480	2,320,642,059	2,322,444,124
Genome Repeat Length, bp	635,420,013	635,976,096	592,124,547	592,611,467	554,061,679	554,491,929
Genome Unique Length, bp	1,649,249,573	1,650,692,901	1,712,180,039	1,713,588,013	1,766,580,380	1,767,952,195
Model Fit, %	96.3584	98.5872	96.6005	98.799	96.8041	98.8033
Read Error Rate, %	0.224321	0.224321	0.219684	0.219684	0.215323	0.215323
Kmergenie						
Property						
Genome Haploid Length (bp)	2,318,146,338					
Predicted Best Kmer	93-95					
SGA Preqc						
Property						
Genome Haploid Length (bp)	2,479,000,000					
Pcr Duplication Proportion	0.06					
Mean quality score	34,25-36					
Fragment Size	400					
Genomescope Reports						
K-mer 27 report:	http://qb.cshl.edu/genomescope/analysis.php?code=3DeFyoNcSdQ7LMiSXsxl					
K-mer 29 report:	http://qb.cshl.edu/genomescope/analysis.php?code=7XldQRDCS7aCXE6XZS2P					
K-mer 31 report:	http://qb.cshl.edu/genomescope/analysis.php?code=6lBxKfeSX2UNetW9t9cl					

Supplementary Table 3. MixS descriptors and accession numbers of tissue sample, raw data and draft genome assembly of *Colossoma macropomum*.

Item	Description		
investigation_type	Eukaryote		
project_name	Colossoma Macropomum Genome assembly		
lat_lon	19°55'45.9"S 40°57'42.1"W		
geo_loc_name	Laranja da Terra, Espirito Santo, Brazil		
tissues	Liver		
collection_date	September/2017		
seq_meth	Illumina HiSeq X Ten		
Collector	Ana Lúcia Salaro		
Maturity	Juvenile		
Datasets Generated	Accession Numbers of NCBI		
Genome raw data	PRJNA596735		
Genome Assembly	JAAFGN000000000		