Gene 591 (2016) 34–42

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Regulatory divergence of homeologous Atlantic salmon *elovl5* genes following the salmonid-specific whole-genome duplication

Greta Carmona-Antoñanzas, Xiaozhong Zheng, Douglas R. Tocher, Michael J. Leaver *

Institute of Aquaculture, School of Natural Sciences, University of Stirling, Pathfoot Building, Stirling FK9 4LA, Scotland, UK

ARTICLE INFO

Article history:
Received 30 May 2016
Accepted 28 June 2016
Available online 30 June 2016

Keywords:
Atlantic salmon
DNA transposon
Homeologous genes
Neofunctionalisation
Transposable elements
Whole-genome duplication
Fatty acid biosynthesis

ABSTRACT

Fatty acyl elongase 5 (elovl5) is a critical enzyme in the vertebrate biosynthetic pathway which produces the physiologically essential long-chain polyunsaturated fatty acids (LC-PUFA), docosahexenoic acid (DHA), and eicosapentenoic acid (EPA) from 18 carbon fatty acids precursors. In contrast to most other vertebrates, Atlantic salmon possess two copies of *elovl5* (*elovl5a* and *elovl5b*) as a result of a whole-genome duplication (WGD) which occurred at the base of the salmonid lineage. WGDs have had a major influence on vertebrate evolution, providing extra genetic material, enabling neofunctionalization to accelerate adaptation and speciation. However, little is known about the mechanisms by which such duplicated homeologous genes diverge. Here we show that homeologous Atlantic salmon *elovl5a* and *elovl5b* genes have been asymmetrically colonised by transposon-like elements. Identical locations and identities of insertions are also present in the rainbow trout duplicate *elovl5* genes, but not in the nearest extant representative preduplicated teleost, the northern pike. Both *elovl5* salmon duplicates possessed conserved regulatory elements that promoted Srebp1- and Srebp2-dependent transcription, and differences in the magnitude of Srebp response between promoters could be attributed to a tandem duplication of SRE and NF-Y cofactor binding sites in *elovl5b*. Furthermore, an insertion in the promoter region of *elovl5a* confers responsiveness to Lxr/Rxr transcriptional activation. Our results indicate that most, but not all, transposon mobilisation into *elovl5* genes occurred after the split from the common ancestor of pike and salmon, but before more recent salmonid speciations, and that divergence of *elovl5* regulatory regions have enabled neofuntionalization by promoting differential expression of these homeologous genes.

1. Introduction

Extant salmonids are descendants of a common ancestor whose genome underwent duplication approximately 88 Myr ago (whole-genome duplication, WGD; Berthelot et al., 2014; Macqueen and Johnston, 2014; Lien et al., 2016). In rainbow trout, about half of these genes have been retained as apparently functional duplicates, while 66% of the remaining singletons still appear to have a pseudogenised duplicate. Furthermore, trout homeologous chromosomes still retain a remarkable colinearity and gene order (Berthelot et al., 2014) suggesting that the salmonid WGD was a result of an autotetraploidisation event. The functional and evolutionary consequences of genome duplication are still unclear, although it has been proposed that more ancient duplications at the base of the vertebrates and early in teleost evolution were a source of extra genetic material leading to diversification, innovation, and ultimately, speciation (Ohno, 1970; Scannell et al., 2006; Van de Peer et al., 2009). Genes duplicated by WGS are referred to as "ohnologues" or "homeologues." Detailed functional comparisons of such duplicate genes from salmonids such as Atlantic salmon (*Salmo salar*) might provide some insight into these processes.

As an outcome of the salmonid WGD, Atlantic salmon possesses more copies of genes for long-chain polyunsaturated fatty acid (LC-PUFA) biosynthetic enzymes compared to other fish (Morais et al., 2009; Monroig et al., 2010; Castro et al., 2012). These genes have been studied in detail and belong to the fatty acyl desaturase (i.e. Fads2) and elongase (i.e. Elovl5) gene families (Morais et al., 2009; Jakobsson et al., 2006), responsible for desaturating and elongating 18 carbon polyunsaturated fatty acids (PUFA), linoleic and α-linolenic acids, to the physiologically critical eicosapentaenoic (EPA), arachidonic (ARA), and docosahexaenoic

Abbreviations: Aa, Amino acid; ARA, Arachidonic acid; bp, Base pair; DHA, Docosahexaenoic acid; ELOVL, Fatty-acyl elongase; EPA, Eicosapentaenoic acid; FAD, Fatty-acyl desaturase; FHM, Fathead minnow; LC-PUFA, Long-chain polyunsaturated fatty acids; LTR, Long terminal repeat; LXR, Liver X receptor; LXRE, Liver X receptor response element; Myr, Million years; NF-Y, Nuclear transcription factor Y; PUFA, Polyunsaturated fatty acids; RXR, Retinoic X receptor; SRE, Sterol response element; SREBP, Sterol regulatory element-binding protein; TE, Transposable element; TF, Transcription factor; TSA, Transcriptome shotgun assembly; TSS, Transcription start site; UTR, Untranslated region; WGD, Whole-genome duplication; WGS, Whole-genome shotgun.

* Corresponding author.
E-mail addresses: gec1@stir.ac.uk (G. Carmona-Antoñanzas), xzheng@staffmail.ed.ac.uk (X. Zheng), d.r.tocher@stir.ac.uk (D.R. Tocher), m.j.leaver@stir.ac.uk (M.J. Leaver).

http://dx.doi.org/10.1016/j.gene.2016.06.056
0378-1119/

(DHA) acids. These genes have been hypothesised, based on previous functional analyses, to have functionally diverged and might have thus physiologically enabled Atlantic salmon to thrive in LC-PUFA-poor environments (Leaver et al., 2008; Carmona-Antoñanzas et al., 2013a). Phylogenetic analyses indicated that duplicated Elovl5 LC-PUFA proteins are subject to strong functional constraints as suggested by comparative studies with the closest extant preduplicated genome, northern pike, *Esox lucius* (Carmona-Antoñanzas et al., 2013a). Although both *elovl5* genes are expressed in LC-PUFA biosynthetic tissues, they are regulated differentially in vivo in response to nutritional changes (Morais et al., 2009), and in vitro they exhibit different responses to transcription factors in cellular transfection assays (Carmona-Antoñanzas et al., 2013b). For example, salmon *elovl5a* responded similarly to the major lipid-regulating transcription factors, sterol regulatory element-binding proteins Srebp1 and Srebp2, whereas *elovl5b* displayed a significantly increased response to Srebp2 (Carmona-Antoñanzas et al., 2013b).

Atlantic salmon aquaculture feeds are now formulated with up to 75% terrestrial plant seed oils instead of the marine oils which were historically used to produce finfish diets Leaver et al., 2008). This is because marine oils harvested from industrial fisheries are now in limiting supply. However, plant oils do not contain EPA/DHA which are characteristically enriched in marine oil and the use of these terrestrial dietary ingredients has led to a reduction in the mass percentage of EPA/DHA present in cultured salmon flesh, with potential effects for fish health and nutritional benefit to human consumers (Sprague et al., 2016). Thus, the endogenous EPA/DHA biosynthetic pathway and the mechanisms by which the pathway is regulated in Atlantic salmon is of considerable interest.

The aims of the present study were to determine the gene structure of duplicate *elovl5* LC-PUFA genes in Atlantic salmon, to compare these with northern pike and rainbow trout *elovl5* genes, and to identify the *cis*-regulatory elements in the salmon promoters which confer the differential responses observed previously. By doing so, we hope to gain insight into the mechanisms by which they are regulated and the patterns of functional divergence of these genes since their duplication in salmonids.

2. Materials and methods

2.1. elovl5 gene structure

An Atlantic salmon genomic DNA library was constructed in lambda FIX II (Stratagene, USA, Zheng et al., 2009). The salmon DNA library was screened with full-length cDNA probes of the salmon elovl5 paralogs, elovl5a [GenBank: AY170327] and elovl5b [GenBank: FJ237531]. Inserts of positive recombinant phage were isolated, fragmented by restriction digest and subcloned to plasmids for sequencing. The full putative elongase genomic nucleotide sequences were assembled using SeqMan II 6.1 module of the Lasergene (DNASTAR Inc., USA). Assembled gene sequences were compared to the Atlantic salmon RefSeq genome assembly (NCBI accession PRJNA287919), and complete gene sequences were inferred from alignment and assembly of matching sequence. Similarly, rainbow trout elovl5 gene sequences were retrieved from the WGS genome assembly (NCBI Accession PRJEB4421). The full gene sequence of northern pike *elovl5* was obtained from the RefSeq genome assembly v1.0 (NCBI accession PRJNA268215).

Atlantic salmon, rainbow trout and pike *elovl5* genomic sequences were compared (blastN) to all Atlantic salmon sequences in Genbank-nr and highly repeated regions identified and these repeated regions were then further screened against Repbase (a database of repetitive element consensus sequences in eukaryotic DNA; Jurka et al., 2005), reported salmon transposons (De Boer et al., 2007) and to an in-house curated database of salmonid genomic repeat sequences. Repeats that shared over 80% identity to consensus sequences of putative mobile elements (Bao and Jurka, 2015a, 2015b) and were >300 bp were scored as transposon-like elements. Full-length elovl5 genes excluding transposon-like elements, were aligned using Mulan (Ovcharenko et al., 2005) and MUSCLE (Edgar, 2004) to identify evolutionary conserved regions across paralogous exons and introns.

2.2. Promoter constructs, deletions, and mutations

The regulatory regions of *elovl5a* (−4898 bp relative to ATG initiation codon; GenBank: GU238431.1) and *elovl5b* (−3143 bp relative to initiation codon; GenBank: GU324549.1) were amplified from genomic DNA using a proof-reading enzyme (Pfu DNA Polymerase, Promega, UK) and primers containing suitable restriction sites (Supplementary Table 1) such that the ATG initiation codon of the luciferase gene in pGL4.10, luc2 (Promega) was replaced by the initiation codon for each *elovl5* gene. The upstream limit for the putative promoter sequence was selected on the basis of the presence of a conserved SacI site immediately beyond which no clear homology between the two *elovl5* sequences could be detected. Thus, the tested promoter regions, in addition to upstream untranscribed sequence, contained transcriptional start sites (TSS), an upstream non-coding exon, and an ATG initiation codon residing within the boundary of the second exon. Each promoter construct was sequenced (Sanger ABI 8730xl, GATC Biotech) to confirm sequence identity and purified using anion-exchange purification columns (QIAfilter plasmid midi kit, Qiagen) for high transfection efficiency. The vectors containing the wild-type full-length promoters, pGL4.10-elovl5a and pGL4.10-elovl5b, were termed SEA1 and SEB1, respectively.

To identify the regions involved in transcription, progressive deletions of *elovl5* gene upstream sequences were constructed using the wild-type reporter constructs (SEA1 and SEB1) as template for PCR amplification and primers containing restriction sites specified in Supplementary Table 1. Eight or six deletion constructs were produced from each (elovl5a, SEA2 to SEA9, and elovl5b, SEB2 to SEB7), each containing the start codon, but representing a shorter version.

Once the regions involved in transcriptional regulation were identified based on the results obtained from promoter deletion analysis, specific sites for mutations were selected using the in silico online MATCHTM, PATCH public 1.0 (Matys et al., 2006), and TFSEARCH tools [http://www.cbrc.jp/research/db/TFSEARCH.html]. Before transfection, all clones were purified using the Qiagen Plasmid Midi Kit (Qiagen) for high transfection efficiency, and constructs verified for accuracy by restriction and sequencing (Sanger ABI 8730xl, GATC Biotech).

Site-directed mutations were performed using the QuickChange II site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. This kit utilises oligonucleotide primers containing the desired mutation. The primers (Supplementary Table 1), each complementary to opposite strands of the vector, were extended by PfuUltra HF DNA polymerase at high annealing temperature (72 °C). The generated amplification product consisted of a mutated circular vector containing staggered nicks at the 5′ end of the amplified strand. Following temperature cycling, the product was digested with 10 U of Dnp I endonuclease, specific for methylated DNA, for 3 h at 37 °C to digest the parental DNA template, thus selecting for the mutated vector. The nicked vector was then transformed into *Escherichia coli* competent TOP10 cells according to the manufacturer's instructions (Invitrogen), which repaired the nick as if it were a DNA polymerase error.

2.3. Cellular transfection assays

For luciferase assays, FHM cells were cultured and transfected as described previously (Carmona-Antoñanzas et al., 2013b). To assess effects of Lxr, Rxr, or Srebps on salmon elovl5 gene promoter activity, FHM cells were cotransfected with pGL4.10-elovl5 constructs (wild promoters, deletion, or site-directed mutants) and nSrebp1 (1–470 aa), nSrebp2 (1–459 aa), Lxr (1–462 aa), and/or Rxr (1–438 aa) expression