Accepted refereed manuscript of:

Carmona-Antonanzas G, Zheng X, Tocher DR & Leaver M (2016) Regulatory divergence of homeologous Atlantic salmon elovl5 genes following the salmonid-specific whole genome duplication, *Gene*, 591 (1), pp. 34-42.

DOI: 10.1016/j.gene.2016.06.056

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

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

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

Corresponding author *: E-mail: m.j.leaver@stir.ac.uk, Tel. +44(0) 1786 466995

Greta Carmona-Antoñanzas gec1@stir.ac.uk

Xiaozhong Zheng: xzheng@staffmail.ed.ac.uk

Douglas R. Tocher: d.r.tocher@stir.ac.uk

Michael J. Leaver: m.j.leaver@stir.ac.uk

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.

Key words: Atlantic salmon, DNA transposon, homeologous genes, neofunctionalisation, transposable elements, whole-genome duplication, fatty acid biosynthesis.

Highlights:

- Ohnologous Atlantic salmon elovl5 genes have been differentially colonised by transposons

- Duplicated elovl5 genes have divergent promoters resulting in activation by Srebp and Lxr transcription factors

- Differential regulation by srebp and Lxr can explain the divergent expression patterns of duplicated elovl5 genes *in vivo*

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; Scannel 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 possess 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 (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-Antonanzas 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-Antonanzas 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-Antonanzas 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-Antonanzas 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 aim of the present study was to determine the gene structure of duplicate *elovl5* LC-PUFA genes in Atlantic salmon, 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 (72ºC). 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 10U of Dnp I endonuclease, specific for methylated DNA, for 3 h at 37 ºC 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 plasmids described previously (Carmona-Antoñanzas et al., 2013b). A reference reporter construct (pGL4.75, hRluc/CMV; Promega) encoding Renilla luciferase was also used as an internal control vector to normalise variations in transfection efficiency. Transfection mixtures consisted of 60 ng of pGL4.10-elovl5 reporter construct (empty pGL4.10 vector in controls), 40 ng of expression vector pcDNA3 (empty pcDNA3 vector in controls), 20 ng of pGL4.75 reference plasmid

(Promega) and 1 µl of Polyfect (Qiagen) transfection reagent. Cells cotransfected with Lxr and Rxr were further treated with the Lxr agonist GW3965 (10 µM), or ethanol carrier added 24 h after transfection. Forty eight hours after all transfections, the medium was aspirated, monolayer washed twice with PBS, cells lysed by 10 min incubation in 75 µl per well of 1x Passive Lysis Buffer (Promega), and Firefly and Renilla luciferase activities were quantified as described in Carmona 2013a. Transactivation activities were obtained using VICTOR X Multilabel Plate reader (PerkinElmer, USA) and data was normalised to the Renilla luciferase activities. Data are presented as means of transactivation activities of the triplicate assays.

2.4 Statistical analyses

All data are presented as means ± SE with three replicates per condition. The effects of deletions or mutations on promoter activities were determined by one-way analysis of variance (ANOVA), followed by multiple comparisons using Tukey's post hoc test. A significance of $P \leq 0.05$ (PASWS 18.0, SPSS Inc., USA) was applied to all statistical tests performed.

3. Results

3.1 Structure and organisation of *elovl5* genes

Four lambda phage salmon genomic inserts with regions of high identity to the previously described salmon elongase cDNAs were obtained after screening an Atlantic salmon genomic DNA library constructed in lambda FIX II. These four phage inserts resolved into two distinct *elovl5* genes, which corresponded to the previously cloned and functionally characterised elongase cDNAs, and the genome sequences were termed *elovl5a* [GenBank: GU238431.1] and *elovl5b* [GenBank:

GU324549.1]. The complete gene sequence for *elovl5a* was obtained, but only a partial sequence corresponding to the first five introns and exons deduced from the *elovl5b* cDNA sequence [GenBank: FJ237531] was obtained. Both gene sequences were subsequently compared and extended with data obtained from the Atlantic salmon genome refrence sequence. These full-length salmon *elovl5* genes were then used to extract *elovl5* gene sequences for rainbow trout based on homology and alignments. Both salmon elongase genes had highly similar coding exon structures, being comprised of one 5'UTR non-coding exon and 7 coding exons that shared a high degree of sequence identity (Figure 1).

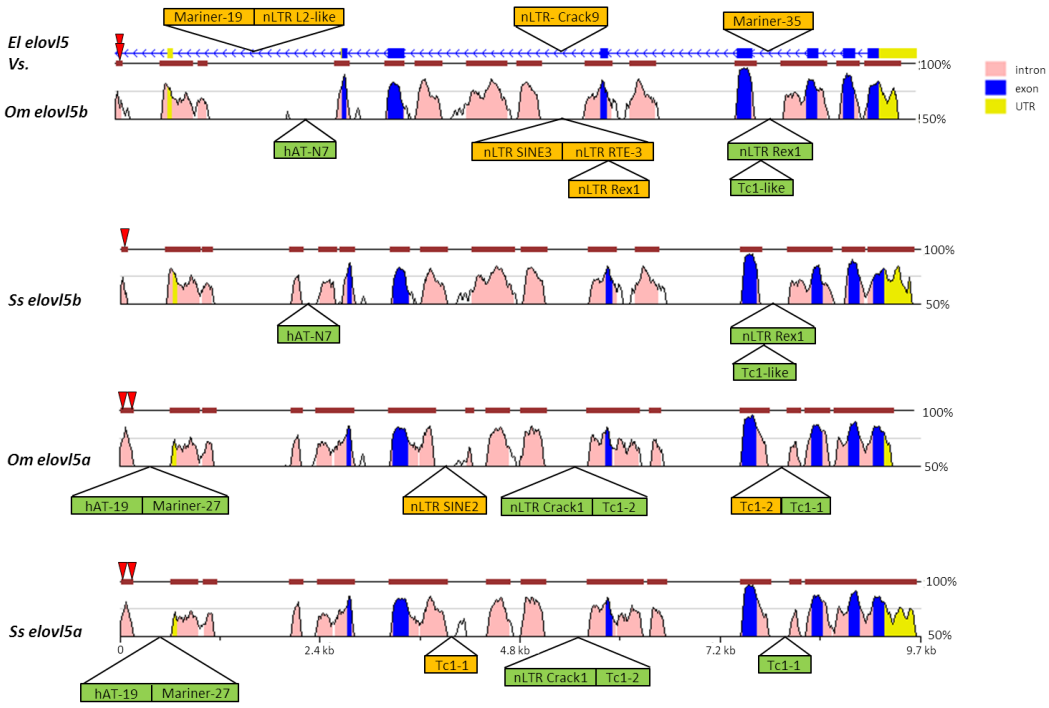


Figure 1. Gene structures of Atlantic salmon, rainbow trout and northern pike *elovl5* homologs. Gene structure was determined from genomic sequences obtained from a genomic DNA lambda FIX II library and complemented with sequences from the NCBI Atlantic salmon, rainbow trout and northern pike whole-genome shotgun contig databases and PCR cloning. Coding exons (blue) predicted from alignments with cDNAs are connected by intron sequences which are plotted as intron regions with > 70% identity over > 100bp in pink in all genes, whereas intron regions with 50 – 70 %

identity over > 100bp are indicated in white The position of predicted non-autonomous DNA transposons and non-LTR (long terminal repeat) retrotransposons that share extensive similarity to RepBase entries from a reference collection of interspersed elements are indicated. Green boxes represent mobile elements that are conserved between orthologs and orange boxes represent non-conserved mobile elements between orthologs. Red triangles indicate Srebp (SRE) binding sites predicted from mutagenesis assays. Ss, *Salmo salar*; Om, *Onycorhynchus mykiss*; El, *Esox lucius*.

A pair of similar rainbow trout genes, and a single pike gene was also identified from the NCBI RefSeq genome databases. Thse genes shared exons of high identity, as well as several areas of clear homology in intron regions and in putative promoter regions. Comparison of salmon and trout *elovl5* genomic sequences to repeat and transposon databases identified several potential mobile genetic elements that were highly repeated throughout both salmon and trout genomes, and which in most cases occurred in the same positions in paralogous salmon and trout *elovl5* genes (Figure 1). Although there were areas of high similarity between pike introns and the salmon and trout introns, transposon-like sequences in the pike gene were different from those in the salmonid genes and were at different locations (Figure 1).

3.2 Atlantic salmon *elovl5* promoter analysis

Sequence homology was assessed between the salmon *elovl5a* and *elovl5b* homeologues including 4.9 kb and 3.1 kb upstream of the initiation codons, respectively. Atlantic salmon *elovl5* homeologues were also compared (BlastN) to the whole-genome shotgun contigs database in Atlantic salmon and Repbase databases to identify intergenic repeated sequences and transposon-like elements present elsewhere in the salmon genome. Highly conserved regions were identified between the *elovl5* duplicates covering ~1.3 kb of the promoters and including the

first exon, which shared 83 % identity in both *elovl5* genes and included the majority of the 5'UTR region. Highly repeated elements were identified in each *elovl5* independently when a genome BlastN produced more than 30 hits throughout the salmon genome and shared > 80 % identity. Highly repeated elements accounted for 23 % of the *elovl5b* promoter and made up 39 % of the *elovl5a* promoter length. The greater size of the *elovl5a* promoter appears to have originated from a 2.5 kb insertion scored as a highly repeated (> 85 % nucleotide identity) sequence scattered throughout the salmon genome and, within this region, signs of ancestral transposable elements (TEs) could be identified (Figure 1). RepeatMasker and CENSOR hits and NCBI megablast analysis, indicated the presence of an hAT-like transposon neighbouring a non-autonomous Mariner-like element, sharing 82 % identity with a previously reported Tc1-like transposase pseudogene, Tcb2 [GenBank: BT059074] from the Atlantic salmon Tc1/Mariner subfamily. Regions that were dissimilar between the two *elovl5* genes and were not highly repeated in the salmon genome accounted for 35 % and 34 % of the total promoter size of the salmon *elovl5a* and *elovl5b*, respectively.

To determine the regions of each salmon *elovl5* gene responsible for transcription factor (TF)-dependent transcription, the DNA fragments encompassing upstream, untranscribed exon and initiation codon sequence, and deletions thereof, was fused to a promoterless luciferase reporter gene (pGL4.10) and cotransfected to fathead minnow FHM cells with constitutive expression vectors encoding the Atlantic salmon transcription factors of interest. Liver X receptor (LXR) forms obligate heterodimers with retinoid X receptor (RXR) and requires ligand activation, and has previously been implicated in salmon elovl5 gene regulation (Carmona-

Antonanzas et al., 2013b) Thus, to test the transcriptional role of Atlantic salmon Lxr, simultaneous cotransfections with piscine Rxra were performed in the presence of the synthetic LXR agonist, GW3965.

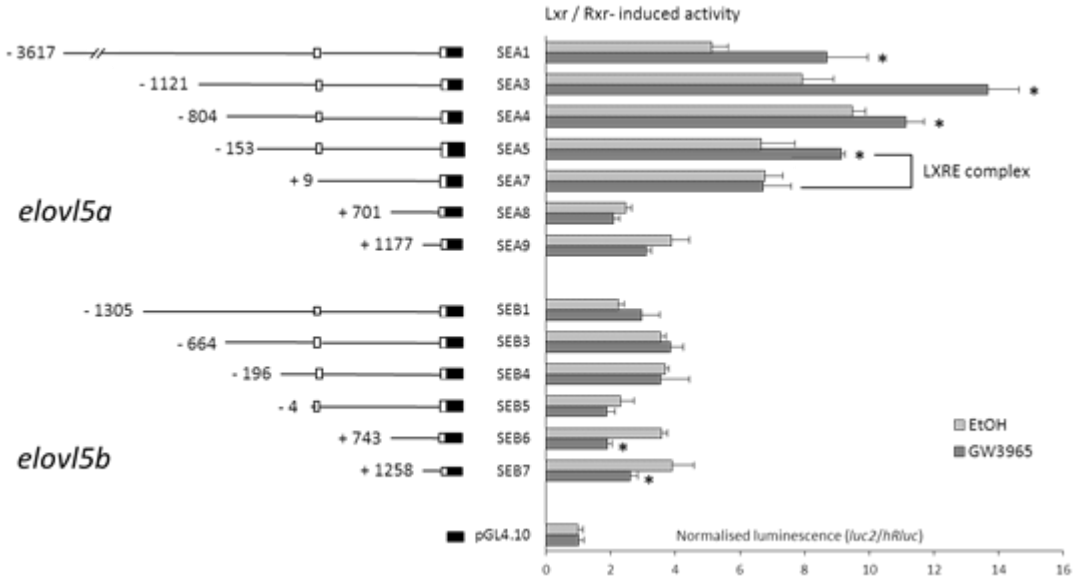


Figure 2. Deletion analyses of salmon *elovl5a* and *elovl5b* gene promoters in the presence of overexpressed salmon Lxr and plaice Rxr transcription factors. Deletion constructs are represented on the left. Non-coding exon is indicated with open boxes and Firefly luciferase coding-sequence by closed boxes. Sequences are numbered relative to the first base of the transcription start site, assumed to be the first base of the 5' non-coding exon. Promoter activity of constructs is represented on the right with the values representing normalised Firefly activity (luc2) to Renilla activity (hRluc). *Indicate the effect of GW3965-activated LXR/RXR is significant compared to ethanol (EtOH) carrier on the same construct (One-way ANOVA; $P < 0.05$). The results are representative of three independent experiments.

Results (Figure 2) were normalised using the *Renilla* luciferase (pGL4.75) and negative controls transfected with promoterless pGL4.10 and mock expression vector (pcDNA3) lacking the insert. Deletion constructs co-tranfected with ligand-activated Lxr and Rxr showed significant differences between *elovl5a* and *elovl5b* responses. Lxr/Rxr stimulation of *elovl5a* promoter, activity was observed when up to 153 nt

upstream of the transcriptional start site (TSS) were included in the construct (SEA5). Constructs SEA1 and SEB1 are the wild-type reporter constructs for *elovl5a* and *elovl5b*, respectively, whereas SEA2-SEA9 and SEB2-SEB7 constitute reporter deletions. Upstream deletion constructs SEA1, SEA3 and SEA4 also showed significant differences, with maximal activity observed when up to 1121 nt upstream of the TSS were included in the construct. In contrast, all deletion constructs excluding the promoter region + 9 upstream of the TSS (SEA7, SEA8 and SEA9) exhibited much reduced activity and no significant differences between the expression levels of the GW3965-activated TF and the mock activated TF with ethanol (Figure 2). These results suggested that *elovl5a* likely possessed an LXR response element (LXRE) within a fragment of 162 nt, located between 153 nt upstream of the TSS and 9 nt downstream the TSS. In contrast, the *elovl5b* promoter was not stimulated by Lxr/Rxr and GW3965 and a significant downregulation, not detected in *elovl5a*, was observed at deletions beyond +743 (Figure 2). These results are indicative of functional divergence of the regulatory regions in the Atlantic salmon *elovl5* homeologues.

3.3 Regulation of salmon *elovl5* duplicates by Srebps

All the deletion constructs from both *elovl5* genes were also tested for response to Srebp1 and Srebp2 overexpression. Srebps are highly conserved transcription factors which regulate, in conjunction with other factors such as nuclear transcription factor Y (NF-Y) and specificity protein 1 (Sp1), pathways of cholesterol and fatty acid biosynthesis in vertebrates (Amemiya-Kudo et al., 2002). Maximal Srebp1 and Srebp2 dependent stimulation of luciferase was observed on the largest *elovl5a* (SEA1) and *elovl5b* (SEB1) promoter constructs. The *elovl5a*

promoter showed 3-fold and 2-fold increases stimulated by Srebp2 and Srebp1, respectively, when up to 3617 nucleotides upstream of the TSSwere included in the wild-type reporter construct (SEA1) (Figure 3).

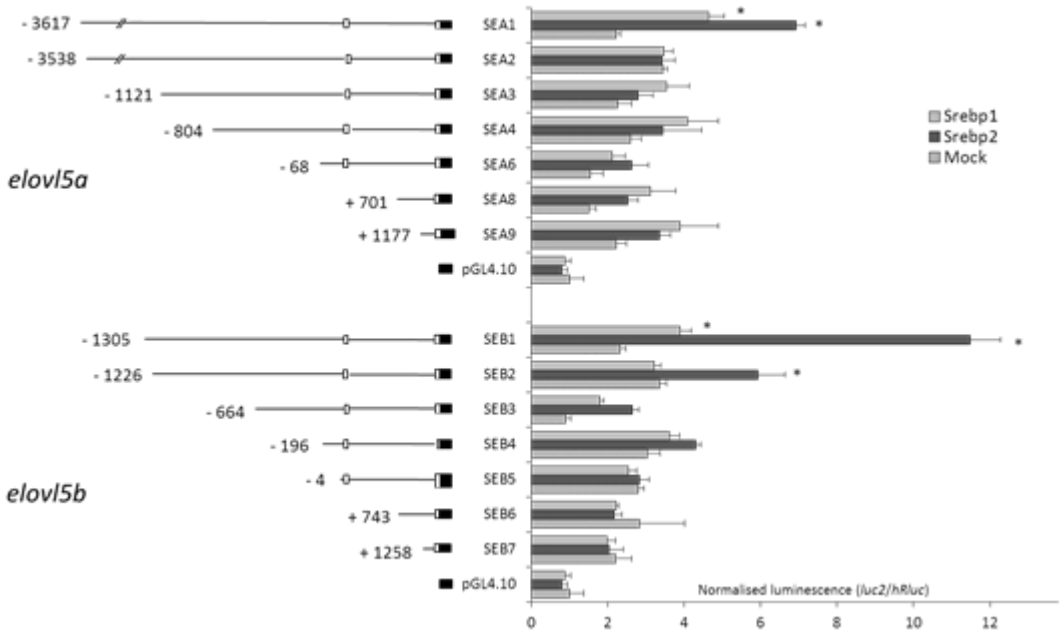


Figure 3. Deletion analyses of salmon *elovl5a* and *elovl5b* gene promoters in the presence of overexpressed nuclear Srebp factors. Deletion constructs are represented on the left. Non-coding exon is indicated with open boxes and Firefly luciferase coding-sequence by closed boxes. Sequences are numbered relative to the first base of the transcription start site, assumed to be the first base of the 5' non-coding exon. Promoter activity of constructs is represented on the right with the values representing normalised Firefly activity (luc2) to Renilla activity (hRluc). *Indicate the effect of SREBP1 or SREBP2 is significant compared to a construct containing no insert (mock) cotransfected with on the same reporter construct (ANOVA; $P < 0.05$). The results are representative of three independent experiments.

A deletion with 3538 nucleotides upstream of the TSS (SEA2) abolished the Srebp1 and Srebp2-dependent stimulation. The *elovl5b* promoter exhibited the highest activity in the presence of Srebp2 (5-fold induction), 3-fold higher than that of Srebp1 and twice as high as the maximum Srebp2-induced activity observed for

335 *elovl5a* (SEA1) (Figure 3) confirming results obtained earlier (Carmona-Antoñanzas
336 at al., 2013b). A deletion, less than 80 nucleotides shorter, including 1226
337 nucleotides upstream of the TSS (SEB2) displayed around 50 % reduction in Srebp2
338 dependent activity, and the Srebp1 dependent effect disappeared in the absence of
339 Srebp. No significant differences in activity were observed in further *elovl5b*
340 promoter deletions. The promoterless pGL4.10 (negative control) exhibited a much-
341 reduced activity in all analyses. These results indicated that Srebp1 and Srebp2 are
342 capable of binding *elovl5a* promoter within a 79 nt region located between 3617 nt
343 (SEA1) and 3538 nt (SEA2) upstream of the TSS. The salmon *elovl5b* promoter
344 responded to Srebp1 within a fragment of 79 nt between 1305 nt (SEB1) and 1226 nt
345 (SEB2) upstream of the TSS, whereas a sterol response element (SRE) responsible
346 solely for Srebp2 regulation appeared to be located within a larger region of 641 nt
347 located between 1305 nt (SEB1) and 664 nt (SEB2) upstream of the TSS.

348 3.4 Promoter mutagenesis analysis

349 Candidate transcriptional regulatory regions in the salmon *elovl5* duplicates,
350 identified by promoter deletions as containing potential Lxr response elements
351 (LXRE) (Figure 2,4) or sterol response elements (SRE) (Figure 3,5), were subjected
352 to *in silico* analysis (Table 1) to identify potential TF binding elements.

Salmon *elovl5a* promoter

```
                                    aa        t     a
-160  GACTGAAGTCATAACTTCATGACCTTTGATCACATGGT  -123
                                SITE 3      SITE 4

                 a   a             a     a
      TTTTGTCAAGTTCATGCAGTCGTCATCTGTAGTCGGAA  -85
                SITE 5            SITE 6

      ATCGGAACATTTTTTATTCCCACAATGCAACACACTTC  -47
      GCAAGGCGTTGACCGGACAAAAGTGATCTGCTCGACGC  -9
      GCCCTGTGTTTCTCCC                        +8
```

353

Figure 4. Nucleotide sequence of the salmon *elovl5a* promoter recognised by Lxr/Rxr. Numbers are given relative to the first base of the transcription start site (TSS). Exonic region corresponding to the 5' non-coding region is shown in bold letters (+1 to +8). Potential transcriptional binding motifs investigated by mutational analyses are indicated by SITE 3-6, with nucleotide mutations indicated by lower-case letters above the wild-type sequences.

```
                                NF-Y                          SRE
B rerio     -624  GCAGCGCTCTGATTGGTCTGTTGTAACTTCAGATGAATGGGGTGATTTACTTGAA-----
E lucius    -599  GCCGTGATTTGATTGGTCTCCACAGAGGAGA-ATGAATGGGGTGACAACTCGGAA-----
S salarA   -3583  GCAGTGATTTGATTGGTTTCCACTGACGAGATGTGAATGGGGTGACA-CTCGGCA-----
O mykissA  -3518  GCAGTGATTTGATTGGTTTCCACTGACGAGATGTGAATGGGGTGACATCTCGGCA-----
S salarB   -1270  GCAGTGATTTGATTGGTCTCCACTGACGAGAAATGAATGGGGTGACATCGGCTTG-32nt
O mykissB  -1267  GCAGTGATTTGATTGGTCTCCACTGACGAGAAATGAATGGGGTGACATCGGTTTG-32nt

S salarB          GCAGTGATTTGATTGGTCTCCACTGATGCGAAATGAATGGGGTGACATCGGCTTGCTTAC
O mykissB         GCACTGATTTGATTGGTCTCCACTGATGCGAAATGAATGGGGTGACATCCGCTTGCTTAC
```

Figure 5. Alignment of *elovl5* duplicated *cis*-regulatory regions in salmonids and northern pike. The numbers indicate the sequences position relative to the transcriptional start sites based on genome sequence information for elovl5 homologs. Shading indicates that nucleotides at that position are identical. NF-Y and SRE binding sites highlighted. NF-Y, nuclear factor Y binding sites; –, represent gaps.

Table 1 Mutated promoter sequences of Atlantic salmon *elovl5* duplicated genes corresponding to putative transcription factor binding sites.

Gene	Construct name	Mutation sites	Binding factor	Sequence	Mutated sequence	Position from TSS
elovl5a	SEA1 M1	1	NF-Y	GATTGGT	GAccaaT	-3576
	SEA1 M2	2	SREBP	TGGGGTGACA	ATaaGaaGAC	-3549
	SEA4 M1	3	PPARα	TGACCT	TGAaaT	-141
	SEA4 M2	4	SREBP1	CACATG	tACATa	-130
	SEA4 M3	5	LXRα	AGTTCA	AaTTaA	-114
	SEA4 M4	6	SREBP1	CATCTG	aATCTa	-99
elovl5b	SEB1 M1	1	NF-Y	GATTGGT	GAccaaT	-1264
	SEB1 M2	3	NF-Y	GATTGGT	GAccaaT	-1178
	SEB1 M3	1, 3				
	SEB1 M4	2	SREBP	TGGGGTGACA	ATaaGaaGAC	-1237
	SEB1 M5	4	SREBP	TGGGGTGACA	ATaaGaaGAC	-1151
	SEB1 M6	2, 4				

Transcriptional binding sites were identified in *elovl5a* [GU238431.1] and *elovl5b* [GU324549.1] promoter regions using the TRANSFAC® binding site database (Matys *et al*., 2006). TSS, transcription start site.

Then, regulatory activity was investigated in FHM cells transfected with mutants targeting potential response elements (Table 1). For salmon *elovl5a*, mutations at sites 3, 5 and 6 caused a reduction of promoter activity of 16 %, 22 % and 25 % relative to the wild type activity (SEA1), respectively, although it was not supported statistically (Figure 6). Mutations of other sites had no effects on promoter activity.

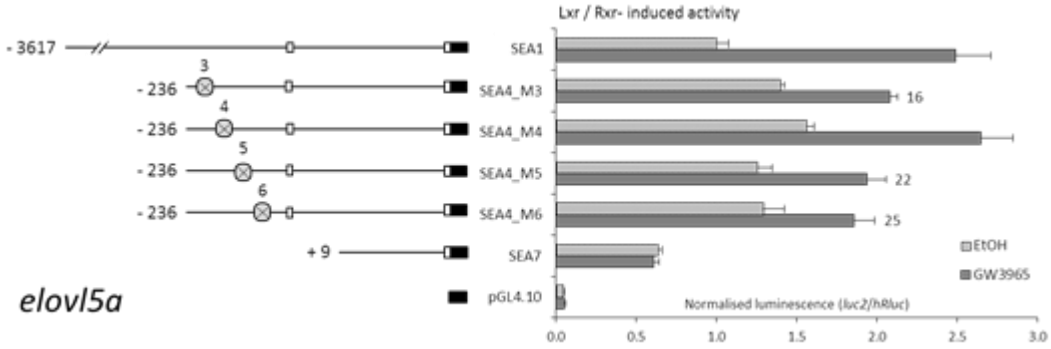
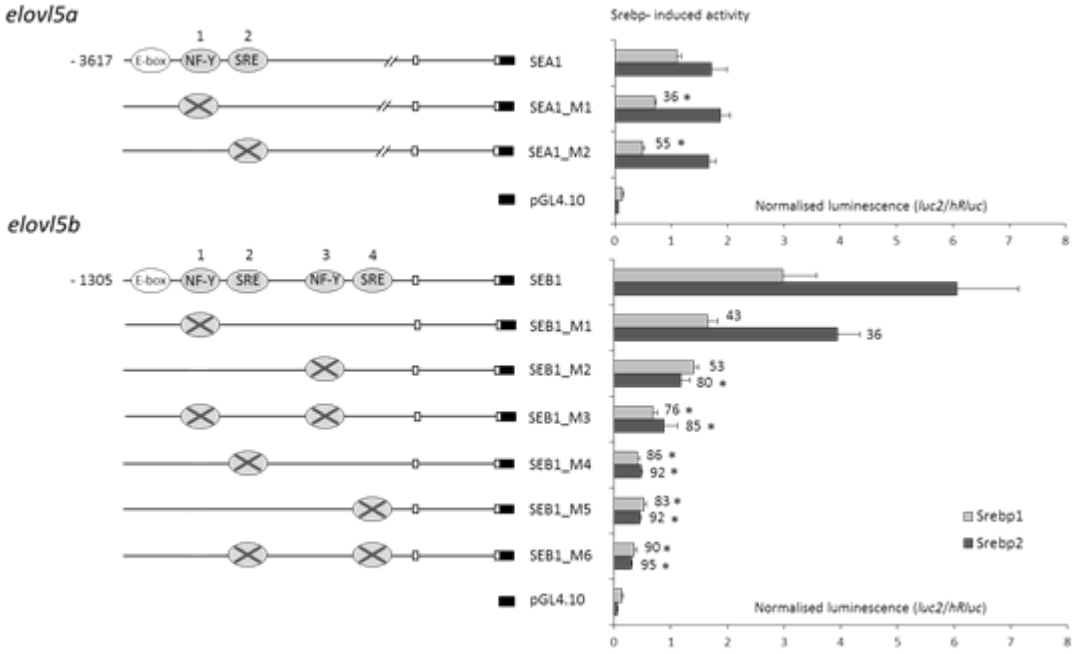


Figure 6. Mutation analyses of salmon *elovl5a* gene promoter in the presence of overexpressed salmon Lxr and plaice Rxr transcription factors. Various mutations of the salmon promoter deletions SEA4 (-236, FIGURE 3) were generated according to FIGURE 5 and the resulting reporter gene activity expressed as the normalised Firefly luminescence (luc2 / hRluc). Relative positions of binding sites, as described in text and in FIGURE 5, are shown and mutation sites within each construct indicated by a closed crossed box. Values represent the percentage expression reduction with respect of the corresponding non-mutated (wild type) promoter of one independent experiment.

Sites identified by in silico analysis as potential NF-Y transcription factor binding and SREs caused significant reductions of *elovl5a* promoter activity of 36 % and 55 % respectively compared to activity of Srebp1 on the wild type sequence (Figure 7).



395

396 **Figure 7. Mutation analyses of salmon *elovl5a* and *elovl5b* gene promoters in the presence of**
397 **overexpressed nuclear Srebp factors.** Various mutations of the salmon promoter deletions SEA1 (-
398 3617, FIGURE 3) and SEB1 (-1305, FIGURE 3) were generated according to FIGURE 6 and the
399 resulting reporter gene activity expressed as the normalised Firefly luminescence (luc2 / hRluc)
400 relative to an expression construct containing no insert. Relative positions of E-box, NF-Y and SRE
401 sites, as described in text and in FIGURE 5, are shown and mutation sites within each construct
402 indicated by a closed crossed box. Values represent the percentage expression reduction with respect
403 of the corresponding non-mutated (wild type) promoter of one independent experiment. *Indicate the
404 effect of Srebp1 or Srebp2 is significant compared to the non-mutated wild-type construct (One-way
405 ANOVA; $P < 0.05$).

406

407 No effect was observed for Srebp2. In *elovl5b*, two NF-Y sites (sites 1 and 3) and
408 two SRE sites (sites 2 and 4) were identified and mutated. For Srebp1-dependent
409 activity on *elovl5b*, independent mutation of NF-Y at sites 1 or 3 caused non-
410 significant reductions of 43 % and 53 % of the wild type activity (SEB1),
411 respectively, whereas the simultaneous mutation of both NF-Y sites (SEB1 M3)
412 caused a significant reduction of 76 % compared to wild type activity. SRE mutation

413 at sites 2 or 4 caused significant reductions of 86 % and 83 %, and simultaneous
414 mutations in sites 2 and 4 reduced the wild type activity by 90 %. Similar results
415 were observed for *elovl5b* promoter mutations when they were co-expressed with
416 Srebp2. Mutations in sites 1, 3, 2 and 4 caused reductions of 36 %, 80 %, 92 % and
417 92 % respectively whereas co-mutation of NF-Y sites 1 and 3, or SRE sites 2 and 4
418 caused the activities to drop 90 % and 95 % of the wild type activity.

419

420 **4. Discussion**

421 This study describes the sequence, structure and regulation of duplicated
422 Atlantic salmon *elovl5* genes and aimed to explain the divergent expression of these
423 genes by comparative analysis of their *cis*-regulatory regions. In addition, by
424 comparing salmon *elovl5a* and *elovl5b* to the rainbow trout and pike homologues we
425 also attempted to reconstruct the evolutionary history of these genes. Salmonids and
426 pike are useful species to study these aspects of evolution since Atlantic salmon and
427 all other salmonids are descended from an ancestor that underwent a WGD event
428 shortly after salmonids and esocids shared a common ancestor (Leong *et al*., 2010,
429 MacQueen and Johnston, 2014). The salmon and trout *elovl5* genes are present in
430 highly syntenic regions, and are most likely duplicated as a consequence of the
431 salmonid WGS (Lien et al., 2016). This indicates, together with phylogenetic
432 evidence (Carmona-Antonanzas et al, 2013a; Lien et al., 2106), that they most likely
433 have arisen as a result of the salmonid whole-genome duplication event timed at
434 about 88 Myr ago. Since WGD is hypothesised to play a major role in evolution,
435 providing a larger complement of ready-made genes on which selection can act, the
436 study of the consequences of WGD might shed light on the processes of

diploidisation after WGS, a requirement for sub- or neofunctionalisation, and ultimately speciation (Wolfe, 2001). Sequence analysis clearly showed that each of the salmon *elovl5* genes had been colonised by a distinct range of transposon-like elements, at distinct sites. Notably the rainbow trout *elovl5a* and *elovl5b* genes are very similar to their Atlantic salmon counterparts, even at the level of transposon insertions. Accordingly, the phylogenetic analysis showed that the salmonid *elovl5a* paralogs clustered together separately from the *elovl5b* paralogs in trout and salmon suggesting they originated after divergence from esocids over 80 Myr ago and prior to trout and salmon speciation, ~ 26 Myr ago (Macqueen and Johnston., 2014). Thus, much of the intron sequence, including the transposon-like elements, is conserved between salmon and trout. Only two repeat sequences were present in salmon *elovl5* genes that were not identifiable in the trout homologues, and these could be categorically identified as Tc1-1, a Tc1/Mariner element (accession, CENSOR). Interestingly other copies of this Tc1-1 element were also present in different introns but in equivalent positions in salmon and trout *elovl5* homologues. The comparison with pike further indicated that *elovl5* introns in this species have been colonised by a different range of transposons, and that the non-repetitive regions of these pike introns can be aligned with the non-repeated, non-transposon-like regions of the corresponding introns in all salmon and trout *elovl5* genes. Some conclusions can be drawn from these observations. Firstly, the asymmetrical distribution pattern of Tc1-like transposons and non-LTR (long terminal repeat) retrotransposons in paralogous *elovl5* genes clearly suggest that Atlantic salmon and rainbow trout *elovl5* genes likely gained most TEs after the species diverged from a common ancestor with the closest pre-duplicated extant relative, northern pike, and after the salmonid *elovl5*

duplication event. Secondly, the peak of activity of the Tc1-1 element might coincide approximately with the divergence of the lineages leading to Atlantic salmon and rainbow trout, since the corresponding *elovl5* genes in these species harbour Tc1-1 both at the same positions, and also in species-specific sites. Rainbow trout and Atlantic salmon genomes show accelerated waves of transposon activity which have been suggested to coincide with speciation (De Boer et al., 2007; Lien et al., 2016; Berthelot et al., 2014), however, it has not been possible to accurately date these mobilisations relative to speciation times due to a lack of knowledge of mutation rates in these elements. The observation that particular transposon activities might overlap with speciation might, given further analysis of duplicated salmonid genes, make it possible to calibrate and derive transposition rates for mobile genetic elements in salmonids, and thus more accurately date waves of their activities. If elevated transposon mobilisation occurred throughout the genome following salmonid polyploidisation it might have played an important role in driving diploidisation, which is required to enable genes duplicated via a WGD to diverge under relaxed selective pressures on at least one of the duplicates.

Furthermore, transposable elements can also contribute promoters and regulatory elements to existing genes (Bejerano et al., 2006; Nishiwara et al., 2006;) and catalyse regulatory divergence of duplicated genes (Herpin eta l., 2010). Genome-scale bioinformatics analyses have shown that many promoters are derived from specific TEs suggesting that insertion of TEs harbouring "ready-to-use" *cis*-regulatory sequences can contribute to the establishment of specific patterns of gene expression (Ferrigno et al., 2001; Mariño-Ramírez et al., 2005). In this regard, Atlantic salmon *elovl5* deletion and mutation constructs allowed us to identify a

region located a few hundred base pairs upstream the transcriptional start site that was stimulated by Lxr agonist GW3965 in the presence of Lxr and obligated heterodimer Rxr. This Lxr/Rxr-dependent response was only detected on one of the duplicates, *elovl5a*, and the sequence element conferring activity is adjacent to an *elovl5a*-specific non-autonomous Mariner-like element. Notably direct LXR-dependent control in LC-PUFA elongation has not been described in vertebrates previously, despite detailed analyses (Wang et al., 2005; Wang et al., 2006; Yoshikawa et al., 2002).

SREBPs, similarly to LXR, are major mediators of lipogenesis,controlled by dietary sterols and fatty acids (Espenshade, 2006). The transfection results clearly showed that Srebp1 and Srebp2 are directly involved in the stimulation of both *elovl5* duplicated Atlantic salmon promoters, which contained sterol response element (SRE)-like sequences in close proximity with NF-Y cofactor sites.. Again these sequence motifs are completely conserved in the rainbow trout *elovl5* genes. In general, Srebp2 stimulated higher response than Srebp1, whereas in mammals the activity of SREBP2 on lipogenic genes is significantly lower than that of SREBP1 (Amemiya-Kudo et al., 2002). Also evident was the differential response of the *elovl5a* and *elovl5b* promoters. As demonstrated by sequence analysis, *elovl5b* possesses adjacent regions with duplicated dyads of SRE and NF-Y binding sites that confer greater SREBP-dependant stimulation than the single dyad in *elovl5a.* It was also apparent that both intact NF-Y and SRE sites were required for maximal Srebp-dependent activity. Furthermore on these particular sites, Srebp2 appeared to promote more transcriptional activity than Srebp1. It is, however, difficult to explain why mutation of either the NF-Y-binding site or the SRE in *elovl5a* did not have any

effect on Srebp-dependent activity, given that the corresponding mutations reduced Srebp1-dependent activity and also both Srebp1- and Srebp2-dependent activity on the *elovl5b* promoter. Overall the transactivation results showed that these *elovl5* promoters were regulated by Srebps through SREs, likely in cooperation with NF-Y. In mammals, SREBPs also require recruitment of NF-Y cofactors to sites adjacent to some SREs (Jackson et al., 1995; Jackson at al., 1998; Näär et al., 1998) for maximal activity.The presence of a single NF-Y and SRE dyad in northern pike suggested that the topology observed in the salmonid *elovl5a* paralogs was likely to be the ancestral state. Zebrafish also contain a very similar element imediately upstream of the single elovl5 gene in this species (Fig 5). This suggests that the SREBP responsive site was duplicated in the *elovl5b* salmonid homeologue after the WGD and prior to the divergence of trout and salmon. Overall, the results indicated that both Lxr/Rxr and Srebp drive different responses between *elovl5* homeologue promoters caused by asymmetrical divergence in *cis*-regulatory regions. This *cis*-regulatory divergence is likely the cause of the previously observed differences in *in vivo* tissue expression patterns and differential responses to nutritional changes in these duplicated genes (Morais et al., 2009; Carmona-Antoñanzas et al., 2013a; Carmona-Antoñanzas et al., 2013b). Unlike many other fish species, Atlantic salmon and related salmonids have the full complement of enzymes, fatty acid elongases and desaturases, required for the biosynthesis of critical LC-PUFA. Here we show that in addition to possessing all of the enzymic machinary for LC-PUFA biosynthesis, duplicated elovl5 genes have also neofunctionalized to enable different regulatory pathways to operate, which, based on *in vivo* observations could increase flexibility in expression across tissues and under different nutritional conditions. Since salmonids spend a large part

of their lives in nutrient-poor freshwaters, these characteristics might represesent adaptations which enable salmonids to more efficiently biosynthesise LC-PUFA, which are in low supply in habitats fuelled predominantly by allochthonous, terrestrial nutrient inputs (Brett and Muller-Navarra, 1997; Leaver et al., 2008)

5. Conclusions

The study of the gene structure of duplicated *elovl5* genes in Atlantic salmon and rainbow trout identified signs of increased transposition following evolutionary divergence with the esocid sister-lineage, but before the most recent speciation events. We suggest that this might have contributed to the formation of stable diploids possibly by inhibiting tetravalent formation and thus enabled duplicate genes to diverge in function, promoting adaptation. Detailed sequence analysis of the target gene promoters presented evidence of asymmetrical distribution of transposon-like elements and divergence of *cis*-regulatory regions in *elovl5a* and *elovl5b*, which resulted in different transactivation responses to transcription factors, LXR and SREBP, involved in the regulation of lipid homeostasis. We obtained evidence of homeologue neofunctionalisation in an *elovl5* duplicate gene possibly associated with a transposon insertion which was responsible for LXR-mediated gene regulatory differences. Also, we detected shared motifs, present at different copy numbers in *elovl5* duplicates in both Atlantic salmon and rainbow trout which conferred response to SREBPs. One consequence of this might be the neofunctionalisation of critical genes of the highly unsaturated fatty biosynthesis pathway, which has enabled salmonids to thrive in nutrient poor freshwater environments.

List of 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.

Competing interests

The authors declare that they have no competing interests.

Author contributions

GCA, DRT and MJL planned and coordinated the research; GCA performed laboratory analyses and data analysis; XZ conducted the gene cloning; GCA wrote the first draft of the manuscript, followed by contributions from remaining authors.

Acknowledgements

This work was supported by a Horizon Scholarship to GEC awarded by the University of Stirling (Scotland).

References

Amemiya-Kudo M, Shimano H, Hasty AH, Yahagi N, Yoshikawa T, Matsuzaka T, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, Osuga J, Harada K, Gotoda T, Sato R, Kimura S, Ishibashi S, Yamada N (2002). Transcriptional activities of nuclear SREBP-1a, -1c, and -2 to different target promoters of lipogenic and cholesterogenic genes. **J Lipid Res** 43, 1220–1235.

Bao W, Jurka J (2015). DNA transposon - a consensus. **Repbase Reports** 15:427.

Bao W, Jurka J (2015). Non-LTR retrotransposon: consensus. **Repbase Reports** 15:685–685.

Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D (2006).A distal enhancer and an ultraconserved exon are derived from a novel retroposon. **Nature** 441, 87–90.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, Aury J-M, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff J-N, Genêt C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y (2014).The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. **Nat Commun** 5, 3657.

Brett MTand Muller-Navarra DC (1997). The role of highly unsaturated fatty acids in aquatic foodweb processes. **Freshwater Biol** 38, 483-499.

Carmona-Antoñanzas G, Tocher DR, Taggart JB, Leaver MJ (2013a). An evolutionary perspective on Elovl5 fatty acid elongase: comparison of Northern pike and duplicated paralogs from Atlantic salmon. **BMC Evol Biol** 13, 85.

Carmona-Antoñanzas G, Tocher DR, Martinez-Rubio L, Leaver MJ (2013b). Conservation of lipid metabolic gene transcriptional regulatory networks in fish and mammals. **Gene** 534, 1–9.

Castro LFC, Monroig Ó, Leaver MJ, Wilson J, Cunha I, Tocher DR (2012). Functional desaturase Fads1 (Δ5) and Fads2 (Δ6) orthologues evolved before the origin of jawed vertebrates. **PLoS One** 7:e31950.

De Boer JG, Yazawa R, Davidson WS, Koop BF (2007). Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. **BMC Genomics** 8, 422.

609 Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high
610 throughput. **Nucleic Acids Res** 32, 1792–7.

611 Espenshade PJ (2006). SREBPs: sterol-regulated transcription factors. **J Cell Sci** 119, 973–
612 976.

613 Ferrigno O, Virolle T, Djabari Z, Ortonne JP, White RJ, Aberdam D (2001).Transposable B2
614 SINE elements can provide mobile RNA polymerase II promoters. **Nat Genet** 28, 77–81.

615 Herpin A, Braasch I, Kraeussling M, Schmidt C, Thoma EC, Nakamura S, Tanaka M, Schartl M
616 (2010). Transcriptional rewiring of the sex determining dmrt1 gene duplicate by
617 transposable elements. **PLos Genet** 6:e1000844.

618 Jackson SM, Ericsson J, Mantovani R, Edwards PA (1998). Synergistic activation of
619 transcription by nuclear factor Y and sterol regulatory element binding protein. **J Lipid Res**
620 39, 767–776.

621 Jackson SM, Ericsson J, Osborne TF, Edwards PA (1995). NF-Y has a novel role in sterol-
622 dependent transcription of two cholesterogenic genes. **J Biol Chem** 270, 21445–8.

623 Jakobsson A, Westerberg R, Jacobsson A (2006). Fatty acid elongases in mammals: their
624 regulation and roles in metabolism. **Prog Lipid Res** 45, 237–249.

625 Jurka J, Kapitonov V V, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J(2005). Repbase
626 Update, a database of eukaryotic repetitive elements. **Cytogenet Genome Res** 110, 462–7.

627 Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR,
628 Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K,
629 Rondeau EB, Di Genova A, Samy JK, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S,
630 Inge Våge D, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ,
631 Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones
632 SJ, Jonassen I, Maass A, Omholt SW, Davidson WS. (2016) The Atlantic salmon genome
633 provides insights into rediploidization. **Nature** 533, 200-205.

634 Leaver, M.J., Bautista, J.M., Björnsson, B.T., Jönsson, E., Krey, G., Tocher, D.R., and
635 Torstensen, B.E. (2008). Towards fish lipid nutrigenomics: current state and prospects for
636 fin-fish aquaculture. **Rev Fish Sci** 16 Supplement 1:73-94.

637 Leong JS, Jantzen SG, von Schalburg KR, Cooper GA, Messmer AM, Liao NY, Munro S, Moore
638 R, Holt RA, Jones SJ, Davidson WS, Koop BF (2010). Salmo salar and Esox lucius full-length
639 cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization
640 genome. **BMC Genomics** 11, 279.

641 Macqueen DJ, Johnston IA (2014). A well-constrained estimate for the timing of the
642 salmonid whole genome duplication reveals major decoupling from species diversification.
643 **Proc R Soc** 281, 20132881.

644 Mariño-Ramírez L, Lewis KC, Landsman D, Jordan IK (2005).Transposable elements donate
645 lineage-specific regulatory sequences to host genomes. **Cytogenet Genome Res** 110, 333–
646 41.

647 Matys V, Kel-Margoulis O V, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev
648 D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE,
649 Wingender E (2006) TRANSFAC and its module TRANSCompel: transcriptional gene
650 regulation in eukaryotes. **Nucleic Acids Res** 34, D108–10.

651 Monroig O, Zheng X, Morais S, Leaver MJ, Taggart JB, Tocher DR (2010).Multiple genes for
652 functional 6 fatty acyl desaturases (Fad) in Atlantic salmon (Salmo salar L.): gene and cDNA
653 characterization, functional expression, tissue distribution and nutritional regulation.
654 **Biochim Biophys Acta** 1801, 1072–1081.

655 Morais S, Monroig O, Zheng XZ, Leaver MJ, Tocher DR (2009). Highly unsaturated fatty acid
656 synthesis in Atlantic salmon: Characterization of ELOVL5-and ELOVL2-like elongases. **Mar**
657 **Biotechnol** 11, 627–639.

658 Näär AM, Beaurang PA, Robinson KM, Oliner JD, Avizonis D, Scheek S, Zwicker J, Kadonaga
659 JT, Tjian R (1998). Chromatin, TAFs, and a novel multiprotein coactivator are required for
660 synergistic activation by Sp1 and SREBP-1a in vitro. **Genes Dev** 12, 3020–3031.

661 Nishihara H, Smit AFA, Okada N (2006).Functional noncoding sequences derived from SINEs
662 in the mammalian genome. **Genome Res** 16, 864–74.

663 Ohno S (1970) Evolution by Gene Duplication. Berlin: Springer-Verlag..

664 Ovcharenko I, Loots GG, Giardine BM, Hou M, Ma J, Hardison RC, Stubbs L, Miller W (2005).
665 Mulan: multiple-sequence local alignment and visualization for studying function and
666 evolution. **Genome** Res 15, 184–94.

667 Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006). Multiple rounds of speciation
668 associated with reciprocal gene loss in polyploid yeasts. **Nature** 440, 341–345.

669 Sprague M, Dick JR, Tocher, DR. (2016). Impact of sustainable feeds on omega-3 long-chain
670 fatty acid levels in farmed Atlantic salmon, 2006–2015. **Scientific Reports**, Article number:
671 21892

672 Taylor JS, Van de Peer Y, Meyer A: Genome duplication, divergent resolution and
673 speciation. Trends Genet 2001, 17:299–301.

674 Van de Peer Y, Maere S, Meyer A (2009. The evolutionary significance of ancient genome
675 duplications. **Nat Rev Genet** 10,725–32.

676 Wang Y, Botolin D, Christian B, Busik J, Xu J, Jump DB (2005).Tissue-specific, nutritional, and
677 developmental regulation of rat fatty acid elongases. **J Lipid Res** 46, 706–715.

678 Wang Y, Botolin D, Xu J, Christian B, Mitchell E, Jayaprakasam B, Nair MG, Peters JM, Busik J
679 V, Olson LK, Jump DB (2006). Regulation of hepatic fatty acid elongase and desaturase
680 expression in diabetes and obesity. **J Lipid Res** 47, 2028–2041.

681 Wolfe KH (2001). Yesterday's polyploids and the mystery of diploidization. **Nat Rev Genet** 2,
682 333-41.

683 Yoshikawa T, Shimano H, Yahagi N, Ide T, Amemiya-Kudo M, Matsuzaka T, Nakakuki M,
684 Tomita S, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, Takahashi A, Sone H, OsugaJi J, Gotoda T,
685 Ishibashi S, Yamada N (2002). Polyunsaturated fatty acids suppress sterol regulatory
686 element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding
687 to LXR response elements. **J Biol Chem** 277, 1705–1711.

688 Zheng X, Leaver MJ, Tocher DR (2009). Long-chain polyunsaturated fatty acid synthesis in
689 fish: Comparative analysis of Atlantic salmon (Salmo salar L.) and Atlantic cod (Gadus
690 morhua L.) Delta6 fatty acyl desaturase gene promoters. **Comp Biochem Physiol B** 154,
691 255–263.

SUPLEMENTARY TABLE 1 Details of primer pairs (restriction sites for *Sac*I, *Xho*I and *Nco*I and mutated sites underlined) used for the construction of promoter deletions and mutations.

Aim	Construct	Forward primer (5' → 3')	Reverse primer (5' → 3')
Promoter deletions	SEA1	AATGAGCTCAGCTCTGCAAAGCCATGTG	AATCTCGAGTTCTGACCTAAATAGACAGATG
	SEA2	AATGAGCTCTCGGCACATGATCGGCC	
	SEA3	AATGAGCTCGAACACGGTGTGGTTGTCTG	
	SEA4	AATGAGCTCACGATCATACCACGCACAAA	
	SEA5	AATGAGCTCTCCCACAATGCAACACACTT	
	SEA6	AATGAGCTCGTATGCGTTTGGGACTGCTT	
	SEA7	AATGAGCTCCCATACATTTCAGGTCTTTC	
	SEB1	TATGAGCTCTGCTCTGTGTAGCCATG	CTCCATGGCTAACCTGAATAGACAGATG
	SEB2	AATGAGCTCCGGCTTGCTAACTGATTTATTA	
	SEB3	TATGAGCTCACAACTAAGATTTTCACTACT	
	SEB4	TATGAGCTCTGGAAACCTAGTTTAAGTGGG	
	SEB5	TATGAGCTCTGTGTTTCTCCCGCTGTTTCC	
	SEB6	TATGAGCTCTGCGTTTGGGACTGCTTGAGC	
	SEB7	TATGAGCTCAAATCAGGGTTCCCCAACTGG	
Promoter mutations	SEA1_M1	GATGCAGTGATTTGACCAATTTCCACTGACGAG	CTCGTCAGTGGAAATTGGTCAAATCACTGCATC
	SEA1_M2	CGAGATGTGAATAAGAAGACACTCGGCACATG	CATGTGCCGAGTGTCTTTTTATTCACATCTCG
	SEA1_M3	GTCATAACTTCATGAAATTTGATCACATGGT	ACCATGTGATCAAATTTCATGAAGTTATGAC
	SEA1_M4	TTCATGACCTTTGATTACATAGTTTTTGTCAAG	CTTGACAAAAACTATGTAATCAAAGGTCATGAA
	SEA1_M5	CATGGTTTTTGTCAAATTAATGCAGTCGTCATC	GATGACGACTGCATTAATTTGACAAAAACCATG
	SEA1_M6	GTTCATGCAGTCGTCATCTGTAGTCGGAAATCG	CGATTTCCGACTACAGATGACGACTGCATGAAC
	SEB1_M1 / SEB1_M3	GATGCAGTGATTTGACCAATCTCCACTGACGA	TCGTCAGTGGAGATTGGTCAAATCACTGCATC
	SEB1_M3 / SEB1_M2	GATGCAGTGATTTGACCAATCTCCACTGATGC	GCATCAGTGGAGATTGGTCAAATCACTGCATC
	SEB1_M4 / SEB1_M6	CGAGAAATGAATAAGAAGACATCGGCTTGCTAACT	AGTTAGCAAGCCGATGTCTTTTTATTCATTTCTCG
	SEB1_M5 / SEB1_M6	TGCGAAATGAATAAGAAGACATCGGCTTGCTTACC	GGTAAGCAAGCCGATGTCTTTTTATTCATTTCGCA