

Accepted refereed manuscript of:

**Carmona-Antonanzas G, Zheng X, Tocher DR & Leaver M (2016)
Regulatory divergence of homeologous Atlantic salmon *elov15* genes
following the salmonid-specific whole genome duplication, *Gene*, 591 (1),
pp. 34-42.**

DOI: [10.1016/j.gene.2016.06.056](https://doi.org/10.1016/j.gene.2016.06.056)

**© 2016, Elsevier. Licensed under the Creative Commons Attribution-
NonCommercial-NoDerivatives 4.0 International
<http://creativecommons.org/licenses/by-nc-nd/4.0/>**

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

3

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

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

8

9

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

11 Greta Carmona-Antoñanzas gec1@stir.ac.uk

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

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

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

15

16 **Abstract**

17 Fatty acyl elongase 5 (*elovl5*) is a critical enzyme in the vertebrate biosynthetic
18 pathway which produces the physiologically essential long chain polyunsaturated
19 fatty acids (LC-PUFA), docosahexenoic acid (DHA) and eicosapentenoic acid (EPA)
20 from 18 carbon fatty acids precursors. In contrast to most other vertebrates, Atlantic
21 salmon possess two copies of *elovl5* (*elovl5a* and *elovl5b*) as a result of a whole
22 genome duplication (WGD) which occurred at the base of the salmonid lineage.

23 WGDs have had a major influence on vertebrate evolution, providing extra genetic
24 material, enabling neofunctionalization to accelerate adaptation and speciation.
25 However, little is known about the mechanisms by which such duplicated
26 homeologous genes diverge. Here we show that homeologous Atlantic salmon
27 *elovl5a* and *elovl5b* genes have been asymmetrically colonised by transposon-like
28 elements. Identical locations and identities of insertions are also present in the
29 rainbow trout duplicate *elovl5* genes, but not in the nearest extant representative
30 preduplicated teleost, the northern pike. Both *elovl5* salmon duplicates possessed
31 conserved regulatory elements that promoted Srebp1- and Srebp2-dependent
32 transcription, and differences in the magnitude of Srebp response between promoters
33 could be attributed to a tandem duplication of SRE and NF-Y cofactor binding sites
34 in *elovl5b*. Furthermore, an insertion in the promoter region of *elovl5a* confers
35 responsiveness to Lxr/Rxr transcriptional activation. Our results indicate that most,
36 but not all transposon mobilisation into *elovl5* genes occurred after the split from the
37 common ancestor of pike and salmon, but before more recent salmonid speciations,
38 and that divergence of *elovl5* regulatory regions have enabled neofunctionalization by
39 promoting differential expression of these homeologous genes.

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

43 **Highlights:**

- 44 • Ohnologous Atlantic salmon *elovl5* genes have been differentially colonised
45 by transposons

- 46 • Duplicated *elov15* genes have divergent promoters resulting in activation by
47 *Srebp* and *Lxr* transcription factors
- 48 • Differential regulation by *srebp* and *Lxr* can explain the divergent expression
49 patterns of duplicated *elov15* genes *in vivo*

50 **1. Introduction**

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

67 As an outcome of the salmonid WGD, Atlantic salmon possess more copies
68 of genes for long-chain polyunsaturated fatty acid (LC-PUFA) biosynthetic enzymes

69 compared to other fish (Morais et al., 2009, Monroig et al., 2010; Castro et al.,
70 2012). These genes have been studied in detail and belong to the fatty acyl
71 desaturase (i.e. Fads2) and elongase (i.e. Elovl5) gene families (Morais et al, 2009;
72 Jakobsson et al., 2006), responsible for desaturating and elongating 18 carbon
73 polyunsaturated fatty acids (PUFA), linoleic and α -linolenic acids, to the
74 physiologically critical eicosapentaenoic (EPA), arachidonic (ARA) and
75 docosahexaenoic (DHA) acids. These genes have been hypothesised, based on
76 previous functional analyses, to have functionally diverged and might have thus
77 physiologically enabled Atlantic salmon to thrive in LC-PUFA-poor environments
78 (Leaver et al., 2008; Carmona-Antonanzas et al., 2013a). Phylogenetic analyses
79 indicated that duplicated Elovl5 LC-PUFA proteins are subject to strong functional
80 constraints as suggested by comparative studies with the closest extant preduplicated
81 genome, northern pike, *Esox lucius* (Carmona-Antonanzas et al., 2013a) . Although
82 both *elovl5* genes are expressed in LC-PUFA biosynthetic tissues, they are regulated
83 differentially *in vivo* in response to nutritional changes (Morais et al., 2009), and in
84 *vitro* they exhibit different responses to transcription factors in cellular transfection
85 assays (Carmona-Antonanzas et al., 2013b) . For example, salmon *elovl5a* responded
86 similarly to the major lipid regulating transcription factors, sterol regulatory element-
87 binding proteins Srebp1 and Srebp2, whereas *elovl5b* displayed a significantly
88 increased response to Srebp2 (Carmona-Antonanzas et al., 2013b).

89 Atlantic salmon aquaculture feeds are now formulated with up to 75%
90 terrestrial plant seed oils instead of the marine oils which were historically used to
91 produce finfish diets (Leaver et al., 2008). This is because marine oils harvested from
92 industrial fisheries are now in limiting supply. However, plant oils do not contain

93 EPA/DHA which are characteristically enriched in marine oil and the use of these
94 terrestrial dietary ingredients has led to a reduction in the mass percentage of
95 EPA/DHA present in cultured salmon flesh, with potential effects for fish health and
96 nutritional benefit to human consumers (Sprague et al., 2016). Thus, the endogenous
97 EPA/DHA biosynthetic pathway and the mechanisms by which the pathway is
98 regulated in Atlantic salmon is of considerable interest.

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

105 **2. Materials and Methods**

106 **2.1 *elovl5* gene structure**

107 An Atlantic salmon genomic DNA library was constructed in lambda FIX II
108 (Stratagene, USA, Zheng et al., 2009). The salmon DNA library was screened with
109 full-length cDNA probes of the salmon *elovl5* paralogs, *elovl5a* [GenBank:
110 AY170327] and *elovl5b* [GenBank: FJ237531]. Inserts of positive recombinant
111 phage were isolated, fragmented by restriction digest and subcloned to plasmids for
112 sequencing. The full putative elongase genomic nucleotide sequences were
113 assembled using SeqMan II 6.1 module of the Lasergene (DNASTAR Inc., USA).
114 Assembled gene sequences were compared to the Atlantic salmon RefSeq genome
115 assembly (NCBI accession PRJNA287919), and complete gene sequences were

116 inferred from alignment and assembly of matching sequence. Similarly, rainbow
117 trout *elovl5* gene sequences were retrieved from the WGS genome assembly (NCBI
118 Accession PRJEB4421). The full gene sequence of northern pike *elovl5* was obtained
119 from the RefSeq genome assembly v1.0 (NCBI accession PRJNA268215).

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

131 2.2 Promoter constructs, deletions and mutations

132 The regulatory regions of *elovl5a* (-4898 bp relative to ATG initiation codon;
133 GenBank: GU238431.1) and *elovl5b* (-3143 bp relative to initiation codon;
134 GenBank: GU324549.1) were amplified from genomic DNA using a proof-reading
135 enzyme (Pfu DNA Polymerase, Promega, UK) and primers containing suitable
136 restriction sites (Supplementary Table 1) such that the ATG initiation codon of the
137 luciferase gene in pGL4.10, luc2 (Promega) was replaced by the initiation codon for
138 each *elovl5* gene. The upstream limit for the putative promoter sequence was

139 selected on the basis of the presence of a conserved SacI site immediately beyond
140 which no clear homology between the two *elovl5* sequences could be detected. Thus,
141 the tested promoter regions, in addition to upstream untranscribed sequence,
142 contained transcriptional start sites (TSS), an upstream non-coding exon and an ATG
143 initiation codon residing within the boundary of the second exon. Each promoter
144 construct was sequenced (Sanger ABI 8730xl, GATC Biotech) to confirm sequence
145 identity and purified using anion-exchange purification columns (QIAfilter plasmid
146 midi kit, Qiagen) for high transfection efficiency. The vectors containing the wild
147 type full-length promoters, pGL4.10-*elovl5a* and pGL4.10-*elovl5b*, were termed
148 SEA1 and SEB1, respectively.

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

155 Once the regions involved in transcriptional regulation were identified based on the
156 results obtained from promoter deletion analysis, specific sites for mutations were
157 selected using the in silico online MATCHTM, PATCH public 1.0 (Matys et al.,
158 2006) and TFSEARCH tools [<http://www.cbrc.jp/research/db/TFSEARCH.html>].
159 Before transfection, all clones were purified using the Qiagen Plasmid Midi Kit
160 (Qiagen) for high transfection efficiency, and constructs verified for accuracy by
161 restriction and sequencing (Sanger ABI 8730xl, GATC Biotech).

162 Site-directed mutations were performed using the QuickChange II site-directed
163 mutagenesis kit (Stratagene) according to the manufacturer's protocol. This kit
164 utilises oligonucleotide primers containing the desired mutation. The primers
165 (Supplementary Table 1), each complementary to opposite strands of the vector,
166 were extended by PfuUltra HF DNA polymerase at high annealing temperature
167 (72°C). The generated amplification product consisted of a mutated circular vector
168 containing staggered nicks at the 5' end of the amplified strand. Following
169 temperature cycling, the product was digested with 10U of Dnp I endonuclease,
170 specific for methylated DNA, for 3 h at 37 °C to digest the parental DNA template,
171 thus selecting for the mutated vector. The nicked vector was then transformed into
172 Escherichia coli competent TOP10 cells according to the manufacturer's instructions
173 (Invitrogen), which repaired the nick as if it were a DNA polymerase error.

174 2.3 Cellular Transfection assays

175 For luciferase assays, FHM cells were cultured and transfected as described
176 previously (Carmona-Antoñanzas et al., 2013b). To assess effects of Lxr, Rxr or
177 Srebps on salmon elov15 gene promoter activity, FHM cells were cotransfected with
178 pGL4.10-elov15 constructs (wild promoters, deletion or site-directed mutants) and
179 nSrebp1 (1-470 aa), nSrebp2 (1-459 aa), Lxr (1-462 aa) and/or Rxr (1-438 aa)
180 expression plasmids described previously (Carmona-Antoñanzas et al., 2013b). A
181 reference reporter construct (pGL4.75, hRluc/CMV; Promega) encoding Renilla
182 luciferase was also used as an internal control vector to normalise variations in
183 transfection efficiency. Transfection mixtures consisted of 60 ng of pGL4.10-elov15
184 reporter construct (empty pGL4.10 vector in controls), 40 ng of expression vector
185 pcDNA3 (empty pcDNA3 vector in controls), 20 ng of pGL4.75 reference plasmid

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

195 2.4 Statistical analyses

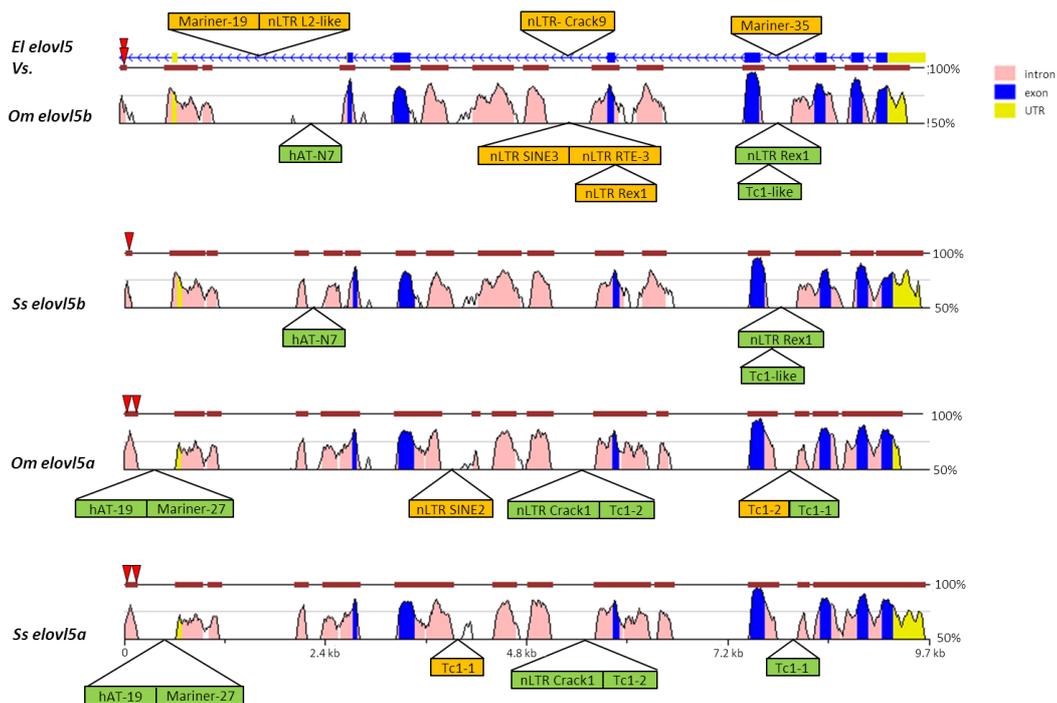
196 All data are presented as means ± SE with three replicates per condition. The effects
197 of deletions or mutations on promoter activities were determined by one-way
198 analysis of variance (ANOVA), followed by multiple comparisons using Tukey's
199 post hoc test. A significance of $P \leq 0.05$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

201 **3. Results**

202 3.1 Structure and organisation of *elovl5* genes

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

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



218

219 **Figure 1. Gene structures of Atlantic salmon, rainbow trout and northern pike *elovl5* homologs.**

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

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

231

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

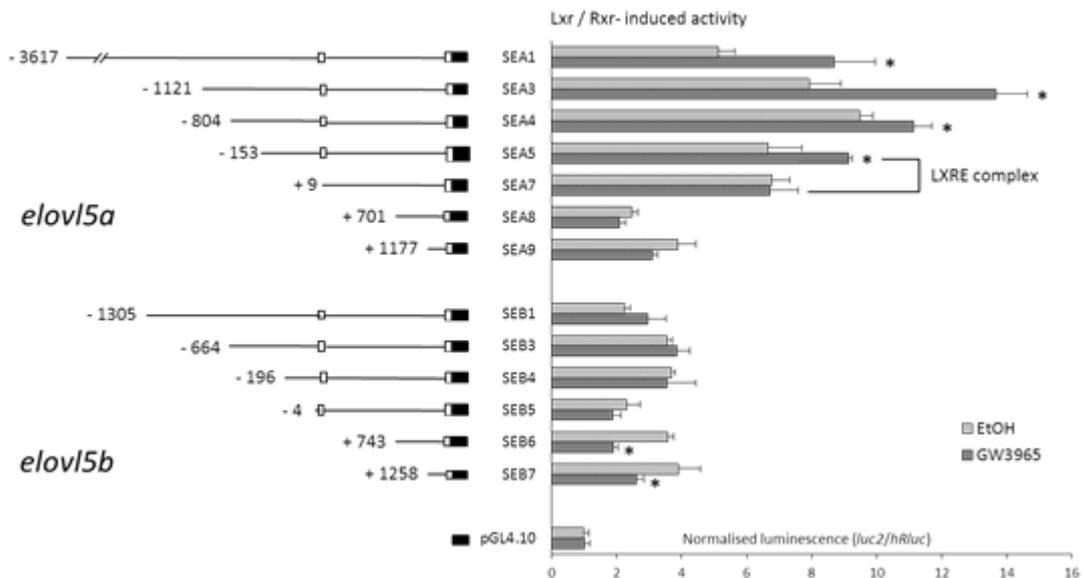
242 3.2 Atlantic salmon *elovl5* promoter analysis

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

250 first exon, which shared 83 % identity in both *elovl5* genes and included the majority
251 of the 5'UTR region. Highly repeated elements were identified in each *elovl5*
252 independently when a genome BlastN produced more than 30 hits throughout the
253 salmon genome and shared > 80 % identity. Highly repeated elements accounted for
254 23 % of the *elovl5b* promoter and made up 39 % of the *elovl5a* promoter length. The
255 greater size of the *elovl5a* promoter appears to have originated from a 2.5 kb
256 insertion scored as a highly repeated (> 85 % nucleotide identity) sequence scattered
257 throughout the salmon genome and, within this region, signs of ancestral
258 transposable elements (TEs) could be identified (Figure 1). RepeatMasker and
259 CENSOR hits and NCBI megablast analysis, indicated the presence of an hAT-like
260 transposon neighbouring a non-autonomous Mariner-like element, sharing 82 %
261 identity with a previously reported Tc1-like transposase pseudogene, Tcb2
262 [GenBank: BT059074] from the Atlantic salmon Tc1/Mariner subfamily. Regions
263 that were dissimilar between the two *elovl5* genes and were not highly repeated in
264 the salmon genome accounted for 35 % and 34 % of the total promoter size of the
265 salmon *elovl5a* and *elovl5b*, respectively.

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

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



277

278 **Figure 2. Deletion analyses of salmon *elov15a* and *elov15b* gene promoters in the presence of**
 279 **overexpressed salmon Lxr and plaice Rxr transcription factors.** Deletion constructs are
 280 represented on the left. Non-coding exon is indicated with open boxes and Firefly luciferase coding-
 281 sequence by closed boxes. Sequences are numbered relative to the first base of the transcription start
 282 site, assumed to be the first base of the 5' non-coding exon. Promoter activity of constructs is
 283 represented on the right with the values representing normalised Firefly activity (luc2) to Renilla
 284 activity (hRluc). *Indicate the effect of GW3965-activated LXR/RXR is significant compared to
 285 ethanol (EtOH) carrier on the same construct (One-way ANOVA; P < 0.05). The results are
 286 representative of three independent experiments.

287

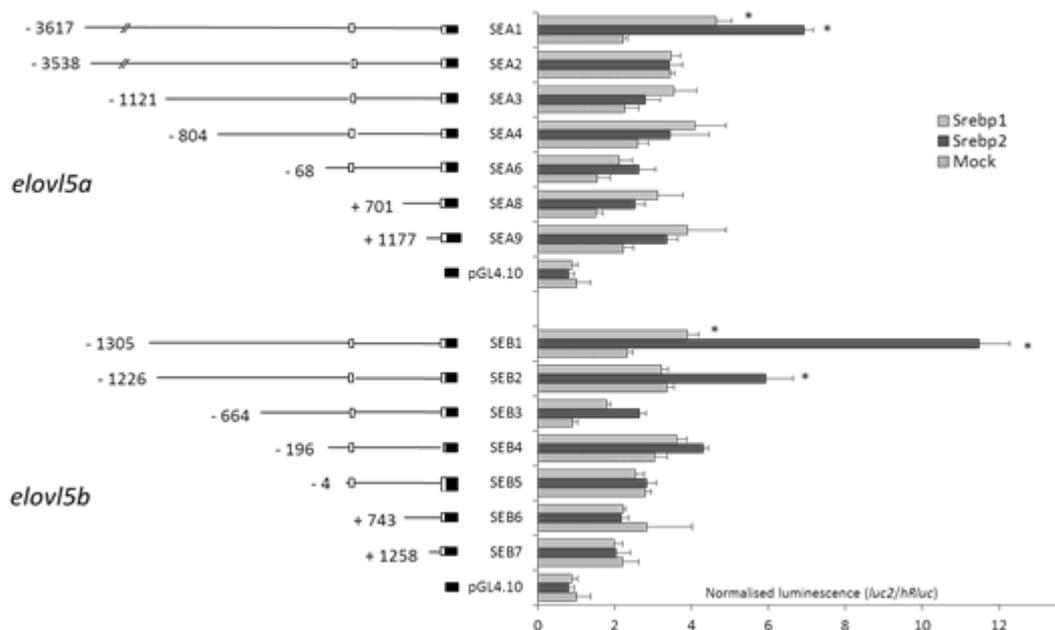
288 Results (Figure 2) were normalised using the *Renilla* luciferase (pGL4.75) and
 289 negative controls transfected with promoterless pGL4.10 and mock expression vector
 290 (pcDNA3) lacking the insert. Deletion constructs co-transfected with ligand-activated
 291 Lxr and Rxr showed significant differences between *elov15a* and *elov15b* responses.
 292 Lxr/Rxr stimulation of *elov15a* promoter, activity was observed when up to 153 nt

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

309 3.3 Regulation of salmon *elovl5* duplicates by Srebps

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

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



320

321 **Figure 3. Deletion analyses of salmon *elovl5a* and *elovl5b* gene promoters in the presence of**
 322 **overexpressed nuclear Srebp factors.** Deletion constructs are represented on the left. Non-coding
 323 exon is indicated with open boxes and Firefly luciferase coding-sequence by closed boxes. Sequences
 324 are numbered relative to the first base of the transcription start site, assumed to be the first base of the
 325 5' non-coding exon. Promoter activity of constructs is represented on the right with the values
 326 representing normalised Firefly activity (luc2) to Renilla activity (hRluc). *Indicate the effect of
 327 SREBP1 or SREBP2 is significant compared to a construct containing no insert (mock) cotransfected
 328 with on the same reporter construct (ANOVA; P < 0.05). The results are representative of three
 329 independent experiments.

330

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

335 *elov15a* (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 *elov15b*
 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 *elov15a* promoter within a 79 nt region located between 3617 nt
 343 (SEA1) and 3538 nt (SEA2) upstream of the TSS. The salmon *elov15b* 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 *elov15* 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 *elov15a* promoter

```

-160 GACTGAAGTCATAACTTTCATGACCTTTGATCACATGGT -123
          aa      t      a
          SITE 3      SITE 4

      a  a      a  a
      SITE 5      SITE 6
TTTTGTCAAGTTCATGCAGTCGTCACTCTGTAGTCGGAA -85

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

353

354 **Figure 4. Nucleotide sequence of the salmon *elov15a* promoter recognised by Lxr/Rxr.** Numbers
 355 are given relative to the first base of the transcription start site (TSS). Exonic region corresponding to
 356 the 5' non-coding region is shown in bold letters (+1 to +8). Potential transcriptional binding motifs
 357 investigated by mutational analyses are indicated by SITE 3-6, with nucleotide mutations indicated by
 358 lower-case letters above the wild-type sequences.

359

			NF-Y		SRE			
360	B rerio	-624	GCAGC	GCTC	TGATTGGTCT	GTTGTA	CTTCAGATGAATGGGGTGA	TTTACTTGAA-----
361	E lucius	-599	GCCGT	GATT	TGATTGGTCT	CCACAG	GGAGA-ATGAATGGGGTGA	CAACTCGGAA-----
362	S salarA	-3583	GCAGT	GATT	TGATTGGTTT	CCACTGA	CGAGATGTGAATGGGGTGA	CA-CTCGGCA-----
363	O mykissA	-3518	GCAGT	GATT	TGATTGGTTT	CCACTGA	CGAGATGTGAATGGGGTGA	CATCTCGGCA-----
364	S salarB	-1270	GCAGT	GATT	TGATTGGTCT	CCACTGA	CGAGAAAATGAATGGGGTGA	CATCGGCTTG-32nt
365	O mykissB	-1267	GCAGT	GATT	TGATTGGTCT	CCACTGA	CGAGAAAATGAATGGGGTGA	CATCGGTTTG-32nt
366								
367	S salarB		GCAGT	GATT	TGATTGGTCT	CCACTGA	TGCGAAAATGAATGGGGTGA	CATCGGCTTGCTTAC
368	O mykissB		GCAGT	GATT	TGATTGGTCT	CCACTGA	TGCGAAAATGAATGGGGTGA	CATCGGCTTGCTTAC
369								

370 **Figure 5. Alignment of *elov15* duplicated *cis*-regulatory regions in salmonids and northern pike.**
 371 The numbers indicate the sequences position relative to the transcriptional start sites based on genome
 372 sequence information for *elov15* homologs. Shading indicates that nucleotides at that position are
 373 identical. NF-Y and SRE binding sites highlighted. NF-Y, nuclear factor Y binding sites; -, represent
 374 gaps.

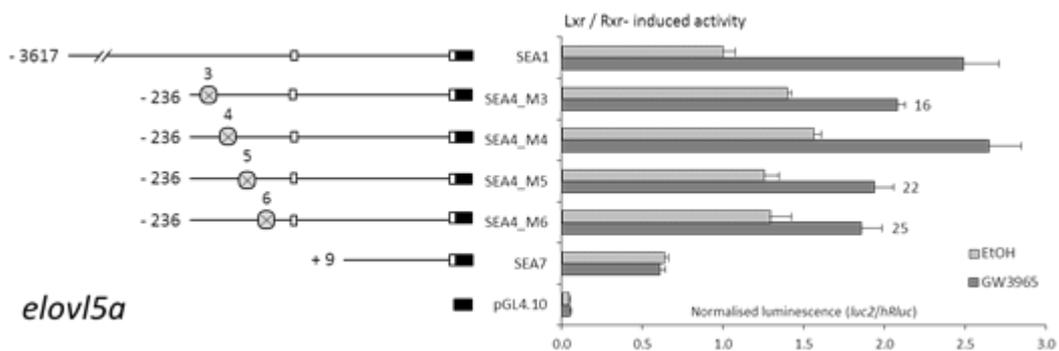
375 **Table 1 Mutated promoter sequences of Atlantic salmon *elov15* duplicated genes corresponding**
 376 **to putative transcription factor binding sites.**

Gene	Construct name	Mutation sites	Binding factor	Sequence	Mutated sequence	Position from TSS
<i>elov15a</i>	SEA1 M1	1	NF-Y	GATTGGT	GA ^c caaT	-3576
	SEA1 M2	2	SREBP	TGGGGTGACA	ATaaGaaGAC	-3549
	SEA4 M1	3	PPAR α	TGACCT	TGAaaT	-141
	SEA4 M2	4	SREBP1	CACATG	tACATa	-130
	SEA4 M3	5	LXR α	AGTTCA	AaTTaA	-114
	SEA4 M4	6	SREBP1	CATCTG	aATCTa	-99
<i>elov15b</i>	SEB1 M1	1	NF-Y	GATTGGT	GA ^c caaT	-1264
	SEB1 M2	3	NF-Y	GATTGGT	GA ^c caaT	-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 *elov15a* [GU238431.1] and *elov15b* [GU324549.1] promoter regions using the TRANSFAC® binding site database (Matys *et al.*, 2006). TSS, transcription start site.

377

378 Then, regulatory activity was investigated in FHM cells transfected with mutants
 379 targeting potential response elements (Table 1). For salmon *elov15a*, mutations at
 380 sites 3, 5 and 6 caused a reduction of promoter activity of 16 %, 22 % and 25 %
 381 relative to the wild type activity (SEA1), respectively, although it was not supported
 382 statistically (Figure 6). Mutations of other sites had no effects on promoter activity.

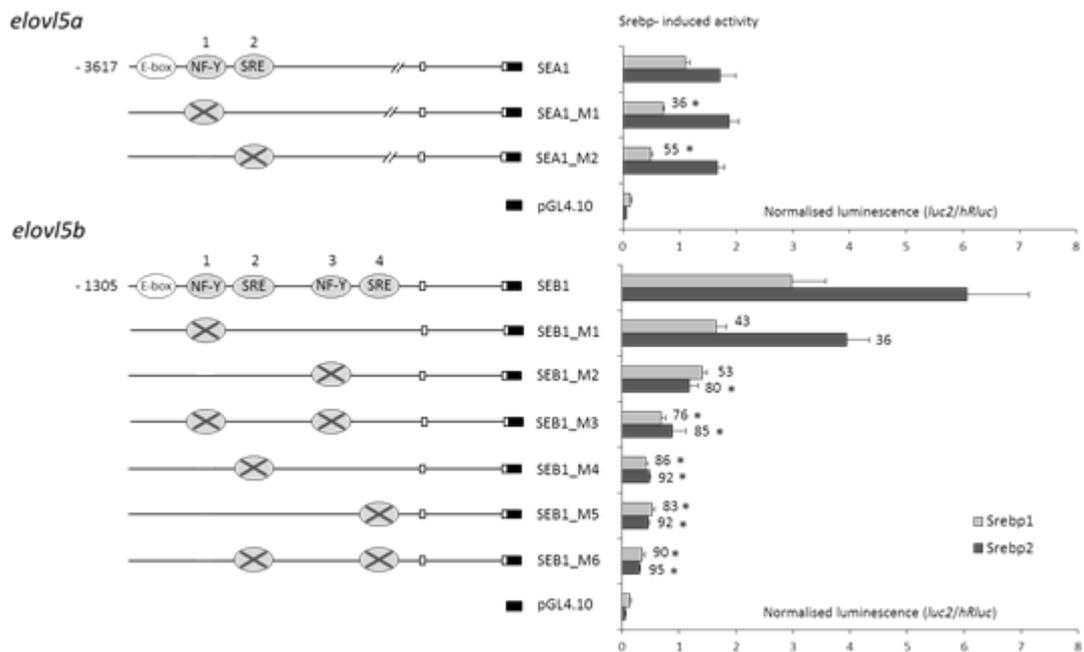


383

384 **Figure 6. Mutation analyses of salmon *elov15a* gene promoter in the presence of overexpressed**
 385 **salmon Lxr and plaice Rxr transcription factors.** Various mutations of the salmon promoter
 386 deletions SEA4 (-236, FIGURE 3) were generated according to FIGURE 5 and the resulting reporter
 387 gene activity expressed as the normalised Firefly luminescence (luc2 / hRluc). Relative positions of
 388 binding sites, as described in text and in FIGURE 5, are shown and mutation sites within each
 389 construct indicated by a closed crossed box. Values represent the percentage expression reduction
 390 with respect of the corresponding non-mutated (wild type) promoter of one independent experiment.

391

392 Sites identified by in silico analysis as potential NF-Y transcription factor binding
 393 and SREs caused significant reductions of *elov15a* promoter activity of 36 % and 55
 394 % respectively compared to activity of Srebp1 on the wild type sequence (Figure 7).



395

396 **Figure 7. Mutation analyses of salmon *elov15a* and *elov15b* 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 *elov15b*, 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 *elov15b*, 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

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

461 duplication event. Secondly, the peak of activity of the Tc1-1 element might coincide
462 approximately with the divergence of the lineages leading to Atlantic salmon and
463 rainbow trout, since the corresponding *elov15* genes in these species harbour Tc1-1
464 both at the same positions, and also in species-specific sites. Rainbow trout and
465 Atlantic salmon genomes show accelerated waves of transposon activity which have
466 been suggested to coincide with speciation (De Boer et al., 2007; Lien et al., 2016;
467 Berthelot et al., 2014), however, it has not been possible to accurately date these
468 mobilisations relative to speciation times due to a lack of knowledge of mutation
469 rates in these elements. The observation that particular transposon activities might
470 overlap with speciation might, given further analysis of duplicated salmonid genes,
471 make it possible to calibrate and derive transposition rates for mobile genetic
472 elements in salmonids, and thus more accurately date waves of their activities. If
473 elevated transposon mobilisation occurred throughout the genome following
474 salmonid polyploidisation it might have played an important role in driving
475 diploidisation, which is required to enable genes duplicated via a WGD to diverge
476 under relaxed selective pressures on at least one of the duplicates.

477 Furthermore, transposable elements can also contribute promoters and
478 regulatory elements to existing genes (Bejerano et al., 2006; Nishiwara et al., 2006;)
479 and catalyse regulatory divergence of duplicated genes (Herpin et al., 2010).

480 Genome-scale bioinformatics analyses have shown that many promoters are derived
481 from specific TEs suggesting that insertion of TEs harbouring “ready-to-use” *cis*-
482 regulatory sequences can contribute to the establishment of specific patterns of gene
483 expression (Ferrigno et al., 2001; Mariño-Ramírez et al., 2005). In this regard,
484 Atlantic salmon *elov15* deletion and mutation constructs allowed us to identify a

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

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

509 effect on Srebp-dependent activity, given that the corresponding mutations reduced
510 Srebp1-dependent activity and also both Srebp1- and Srebp2-dependent activity on
511 the *elovl5b* promoter. Overall the transactivation results showed that these *elovl5*
512 promoters were regulated by Srebps through SREs, likely in cooperation with NF-Y.
513 In mammals, SREBPs also require recruitment of NF-Y cofactors to sites adjacent to
514 some SREs (Jackson et al., 1995; Jackson et al., 1998; Näär et al., 1998) for maximal
515 activity. The presence of a single NF-Y and SRE dyad in northern pike suggested that
516 the topology observed in the salmonid *elovl5a* paralogs was likely to be the ancestral
517 state. Zebrafish also contain a very similar element immediately upstream of the single
518 *elovl5* gene in this species (Fig 5). This suggests that the SREBP responsive site was
519 duplicated in the *elovl5b* salmonid homeologue after the WGD and prior to the
520 divergence of trout and salmon. Overall, the results indicated that both Lxr/Rxr and
521 Srebp drive different responses between *elovl5* homeologue promoters caused by
522 asymmetrical divergence in *cis*-regulatory regions. This *cis*-regulatory divergence is
523 likely the cause of the previously observed differences in *in vivo* tissue expression
524 patterns and differential responses to nutritional changes in these duplicated genes
525 (Morais et al., 2009; Carmona-Antoñanzas et al., 2013a; Carmona-Antoñanzas et al.,
526 2013b). Unlike many other fish species, Atlantic salmon and related salmonids have
527 the full complement of enzymes, fatty acid elongases and desaturases, required for
528 the biosynthesis of critical LC-PUFA. Here we show that in addition to possessing
529 all of the enzymic machinery for LC-PUFA biosynthesis, duplicated *elovl5* genes
530 have also neofunctionalized to enable different regulatory pathways to operate,
531 which, based on *in vivo* observations could increase flexibility in expression across
532 tissues and under different nutritional conditions. Since salmonids spend a large part

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

537 **5. Conclusions**

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

556

557 **List of abbreviations**

558 Aa, amino acid; ARA, arachidonic acid; bp, base pair; DHA, docosahexaenoic acid;
559 ELOVL, fatty-acyl elongase; EPA, eicosapentaenoic acid; FAD, fatty-acyl
560 desaturase; FHM, fathead minnow; LC-PUFA, long-chain polyunsaturated fatty
561 acids; LTR, long terminal repeat; LXR, liver X receptor; LXRE, liver X receptor
562 response element; Myr, million years; NF-Y, nuclear transcription factor Y; PUFA,
563 polyunsaturated fatty acids; RXR, retinoic X receptor; SRE, sterol response element;
564 SREBP, sterol regulatory element-binding protein; TE, transposable element; TF,
565 transcription factor; TSA, transcriptome shotgun assembly; TSS, transcription start
566 site; UTR, untranslated region; WGD, whole genome duplication; WGS, whole-
567 genome shotgun.

568

569 **Competing interests**

570 The authors declare that they have no competing interests.

571

572 **Author contributions**

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

576

577 **Acknowledgements**

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

580 **References**

- 581 Amemiya-Kudo M, Shimano H, Hasty AH, Yahagi N, Yoshikawa T, Matsuzaka T, Okazaki H,
582 Tamura Y, Iizuka Y, Ohashi K, Osuga J, Harada K, Gotoda T, Sato R, Kimura S, Ishibashi S,
583 Yamada N (2002). Transcriptional activities of nuclear SREBP-1a, -1c, and -2 to different
584 target promoters of lipogenic and cholesterologenic genes. **J Lipid Res** 43, 1220–1235.
- 585 Bao W, Jurka J (2015). DNA transposon - a consensus. **Rebase Reports** 15:427.
- 586 Bao W, Jurka J (2015). Non-LTR retrotransposon: consensus. **Rebase Reports** 15:685–685.
- 587 Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D
588 (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon.
589 **Nature** 441, 87–90.
- 590 Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C,
591 Labadie K, Alberti A, Aury J-M, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C,
592 Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff J-N,
593 Genêt C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y (2014). The rainbow trout genome
594 provides novel insights into evolution after whole-genome duplication in vertebrates. **Nat**
595 **Commun** 5, 3657.
- 596 Brett MT and Muller-Navarra DC (1997). The role of highly unsaturated fatty acids in aquatic
597 foodweb processes. **Freshwater Biol** 38, 483-499.
- 598 Carmona-Antoñanzas G, Tocher DR, Taggart JB, Leaver MJ (2013a). An evolutionary
599 perspective on Elovl5 fatty acid elongase: comparison of Northern pike and duplicated
600 paralogs from Atlantic salmon. **BMC Evol Biol** 13, 85.
- 601 Carmona-Antoñanzas G, Tocher DR, Martinez-Rubio L, Leaver MJ (2013b). Conservation of
602 lipid metabolic gene transcriptional regulatory networks in fish and mammals. **Gene** 534,
603 1–9.
- 604 Castro LFC, Monroig Ó, Leaver MJ, Wilson J, Cunha I, Tocher DR (2012). Functional
605 desaturase Fads1 ($\Delta 5$) and Fads2 ($\Delta 6$) orthologues evolved before the origin of jawed
606 vertebrates. **PLoS One** 7:e31950.
- 607 De Boer JG, Yazawa R, Davidson WS, Koop BF (2007). Bursts and horizontal evolution of
608 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 cholesterologenic genes. **J Biol Chem** 270, 21445–8.

623 Jakobsson A, Westerberg R, Jakobsson 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, Osuga 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.

SUPPLEMENTARY TABLE 1 Details of primer pairs (restriction sites for *SacI*, *XhoI* and *NcoI* 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	AAT <u>GAGCTC</u> AGCTCTGCAAAGCCATGTG	AAT <u>CTCGAGT</u> TCTGACCTAAATAGACAGATG
	SEA2	AAT <u>GAGCTC</u> TGGGCACATGATCGGCC	
	SEA3	AAT <u>GAGCTC</u> GAAACACGGTGTGGTTGTCTG	
	SEA4	AAT <u>GAGCTC</u> ACGATCATAACCACGCACAAA	
	SEA5	AAT <u>GAGCTC</u> TCCCACAATGCAACACACTT	
	SEA6	AAT <u>GAGCTC</u> GTATGCGTTTGGGACTGCTT	
	SEA7	AAT <u>GAGCTC</u> CCATACATTTTCAGGTCCTTC	
	SEB1	TAT <u>GAGCTC</u> TGCTCTGTGTAGCCATG	CTCCATGGCTAACCTGAATAGACAGATG
	SEB2	AAT <u>GAGCTC</u> CGGCTTGCTAACTGATTTATTA	
	SEB3	TAT <u>GAGCTC</u> CAACTAAGATTTTCACTACT	
	SEB4	TAT <u>GAGCTC</u> TGGAAACCTAGTTTAAGTGGG	
	SEB5	TAT <u>GAGCTC</u> TGTGTTTCTCCGCTGTTCC	
	SEB6	TAT <u>GAGCTC</u> TGCGTTTGGGACTGCTTGAGC	
	SEB7	TAT <u>GAGCTC</u> AAATCAGGGTTCCTCCAACTGG	
Promoter mutations	SEA1_M1	GATGCAGTGATTT <u>GACCAAT</u> TCCACTGACGAG	CTCGTCAGTGAAAT <u>TGGTCAAAT</u> CACTGCATC
	SEA1_M2	CGAGATGTGAAT <u>AAGAAGAC</u> ACTCGGCACATG	CATGTGCCGAG <u>TGCTTTTT</u> ATTCACATCTCG
	SEA1_M3	GTCATAACTTCA <u>TGAAAT</u> TGATCACATGGT	ACCATGTGATCAAAT <u>TTTCAT</u> GAAGTTATGAC
	SEA1_M4	TTTATGACCTTTGAT <u>TACATAG</u> TTTTGTCAAG	CTTGACAAAAAC <u>TATGTA</u> ATCAAAGGTCATGAA
	SEA1_M5	CATGGTTTTGTCA <u>AAATTA</u> ATGCAAGTCGTCATC	GATGACGACTGCAT <u>TAATTT</u> GACAAAAACCATG
	SEA1_M6	GTTTCATGCAGTCG <u>TATCTG</u> TAGTCGGAAATCG	CGATTTCCGACT <u>ACAGATG</u> ACGACTGCATGAAC
	SEB1_M1 / SEB1_M3	GATGCAGTGATTT <u>GACCAAT</u> TCCACTGACGA	TCGTCAAGTGGAG <u>ATTGGTCAAAT</u> CACTGCATC
	SEB1_M3 / SEB1_M2	GATGCAGTGATTT <u>GACCAAT</u> TCCACTGATGC	GCATCAGTGGAG <u>ATTGGTCAAAT</u> CACTGCATC
	SEB1_M4 / SEB1_M6	CGAGAAATGAAT <u>AAGAAGAC</u> ATCGGCTTGCTAACT	AGTTAGCAAGCCGAT <u>TGCTTTTT</u> ATTCATTTCTCG
	SEB1_M5 / SEB1_M6	TGCGAAATGAAT <u>AAGAAGAC</u> ATCGGCTTGCTTACC	GGTAAGCAAGCCGAT <u>TGCTTTTT</u> ATTCATTTCTCGA