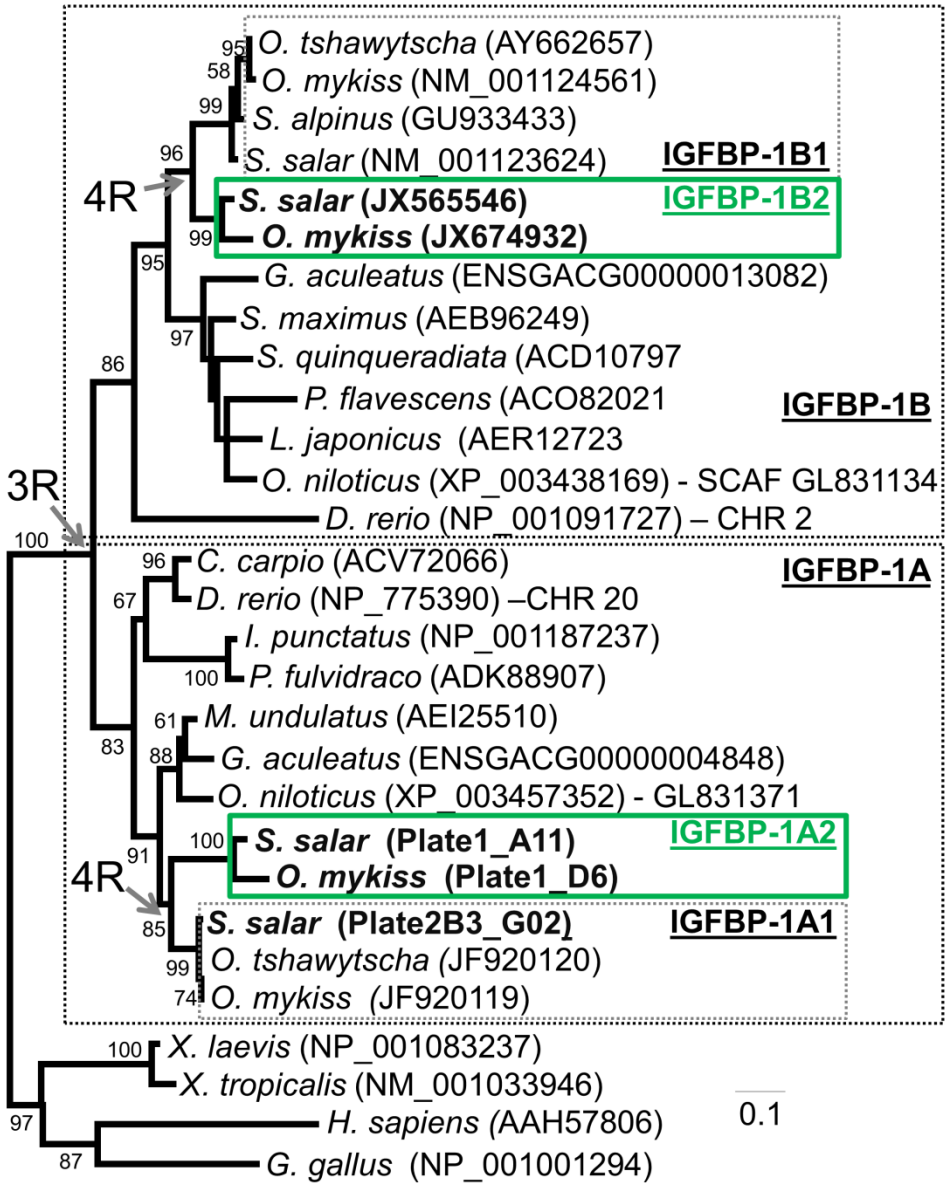

Figure S1. IGFBP-1 NJ tree generated from equivalent data as the ML tree provided/described in the main text, Fig. 1A. The two trees showed very high congruence.

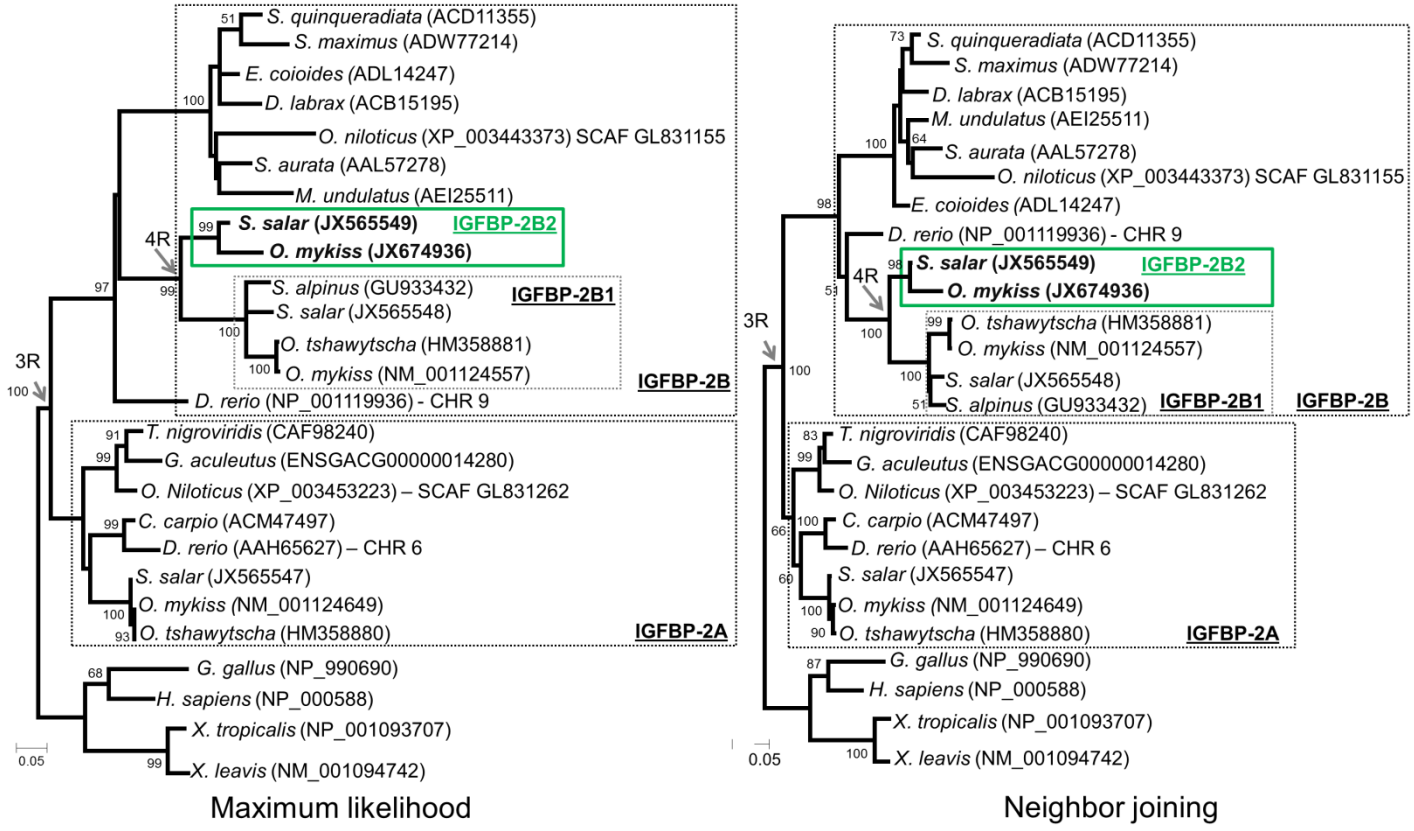

Figure S2. IGFBP-2 family member ML and NJ trees. Both tree-building methods generated largely congruent topologies meeting the criteria for 3R and 4R set out in the main text. However, the branching of teleost sequences did not always follow expected taxonomic relationships, with Ostariophysi sometimes branching as sister to Protacanthopterygi. However, because synteny results fulfilled the criteria set out in the main text (Fig. S6), we suggest that these minor branching patterns are artificial and that IGFBP-2 duplicated during 3R producing IGFBP-2A and -2B (after Zhou et al. 2008), before each co-orthologue duplicated again during 4R, respectively producing IGFBP-2A1, -2A2, -2B1 and 2B2 (Table 1, main text). It is inferred that one IGFBP-2A paralogue was lost in the salmonine ancestor.

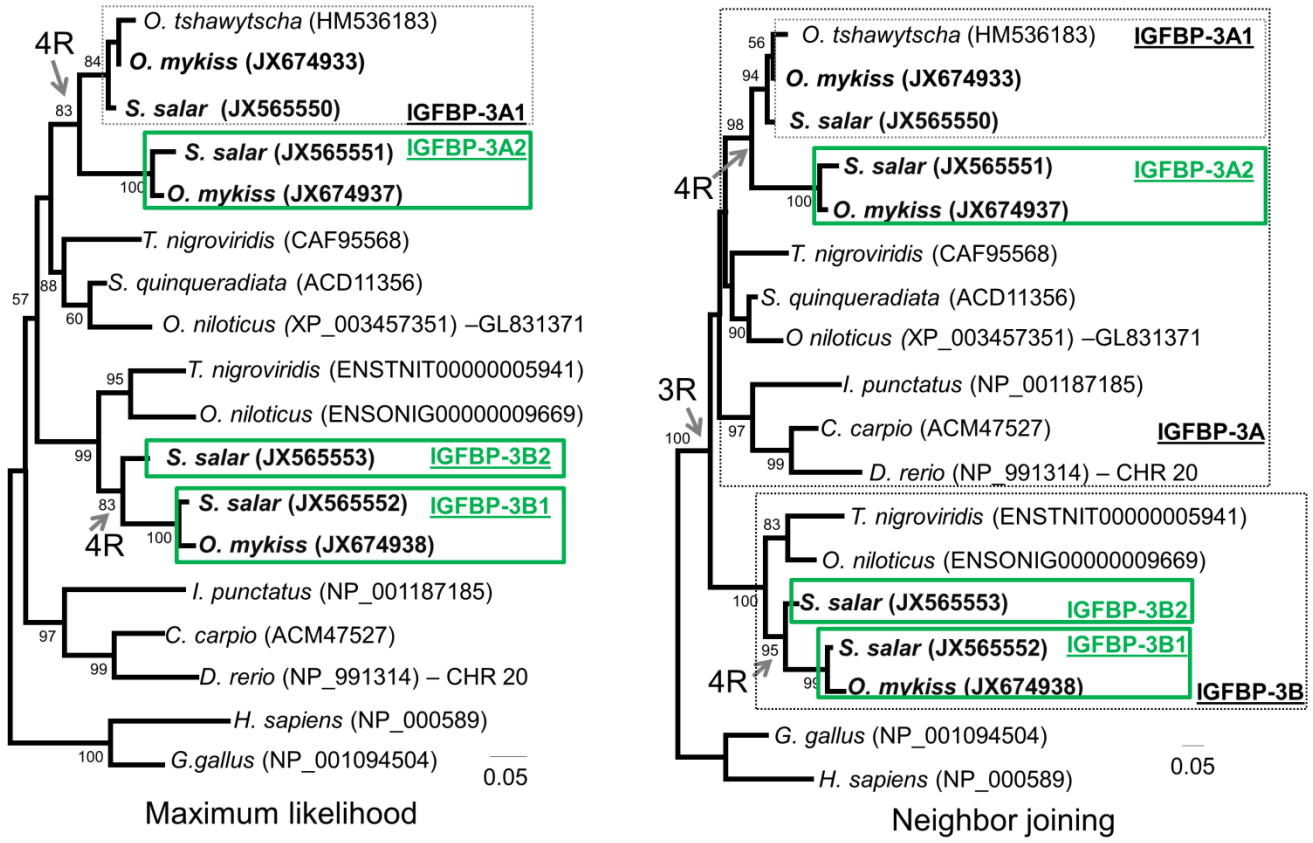

Figure S3. IGFBP-3 family member ML and NJ trees. These tree-building methods did not produce congruent trees, owing to the divergent branching of the Ostariophysii clade, which represents the single IGFBP-3 gene of this lineage (Daza et al. 2011). By ML, this clade is basal to all teleost IGFBP-3 sequences, which, in turn, splits into two clades (57% bootstrap support), each containing Protacanthoptergii and Acanthopterygii sister groups. By NJ, the Ostariophysii forms the basal branch of a teleost-wide clade (<50% bootstrap support), separated from another clade containing Protacanthoptergii and Acanthopterygii sister groups. Thus, NJ meets the criteria for 3R set out in the main text, albeit without statistical support. However, the results from both methods match criteria for 4R set out in the main text and the synteny analyses fulfill the 3R criteria (Fig. S6). Thus, we consider it most parsimonious that IGFBP-3 duplicated during 3R producing IGFBP-3A and -3B (after Shimizu et al. 2011a), before each paralogue-pair duplicated again during 4R, producing IGFBP-3A1, -3A2, -3B1 and -3B2 (Table 1, main text).

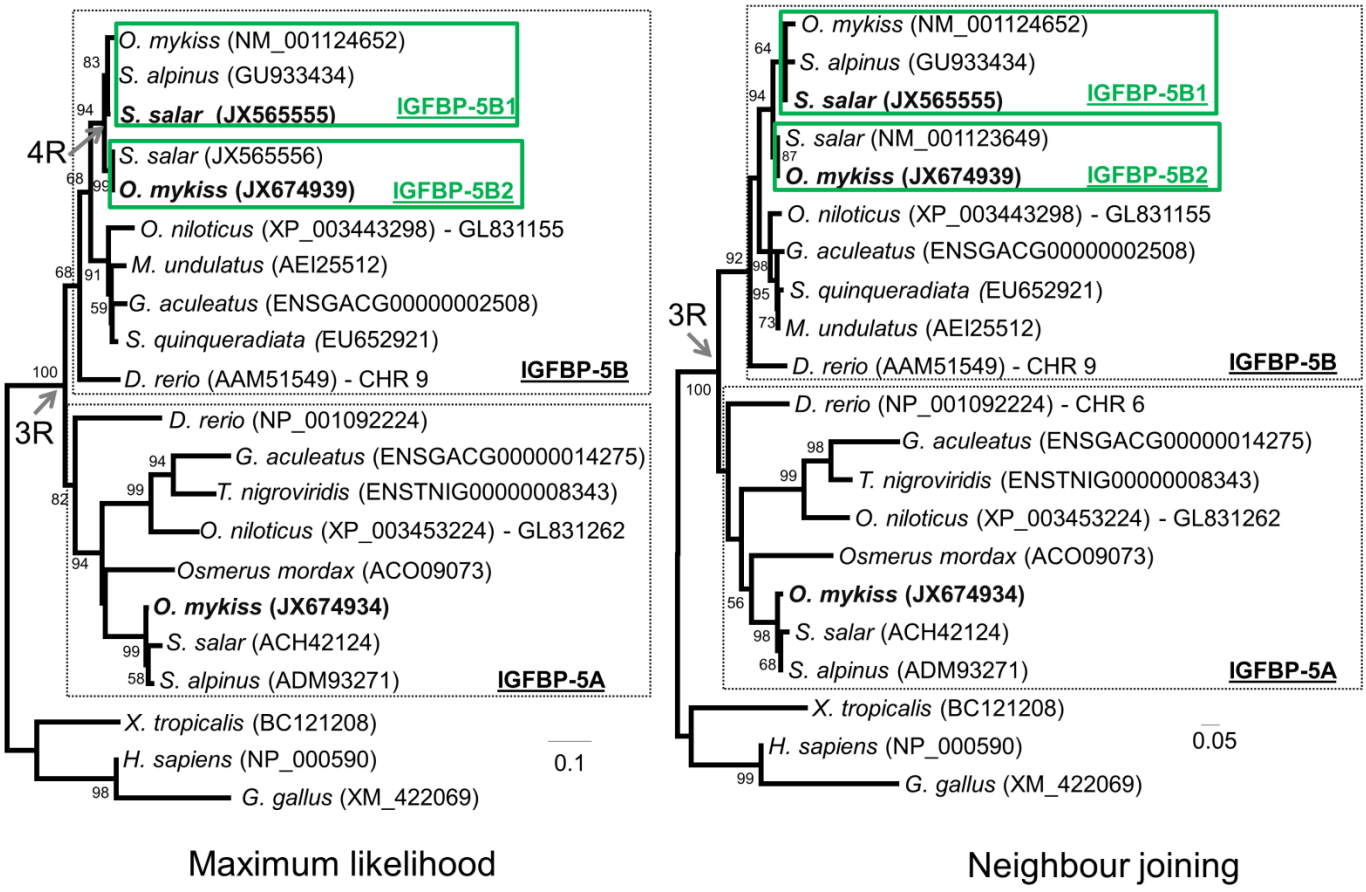

Figure S4. IGFBP-5 family member ML and NJ trees. Both tree building methods and the synteny analysis (Fig. S6) meet the criteria for 3R/4R set out in the main text. This suggests that IGFBP-5 duplicated during 3R producing IGFBP-5A and -5B (after Dai et al. 2010), before each gene duplicated again during 4R, respectively producing IGFBP-5A1, -5A2, -5B1 and -5B2 (Table 1, main text). It is inferred that one IGFBP-5A paralogue was lost in the salmonine ancestor.

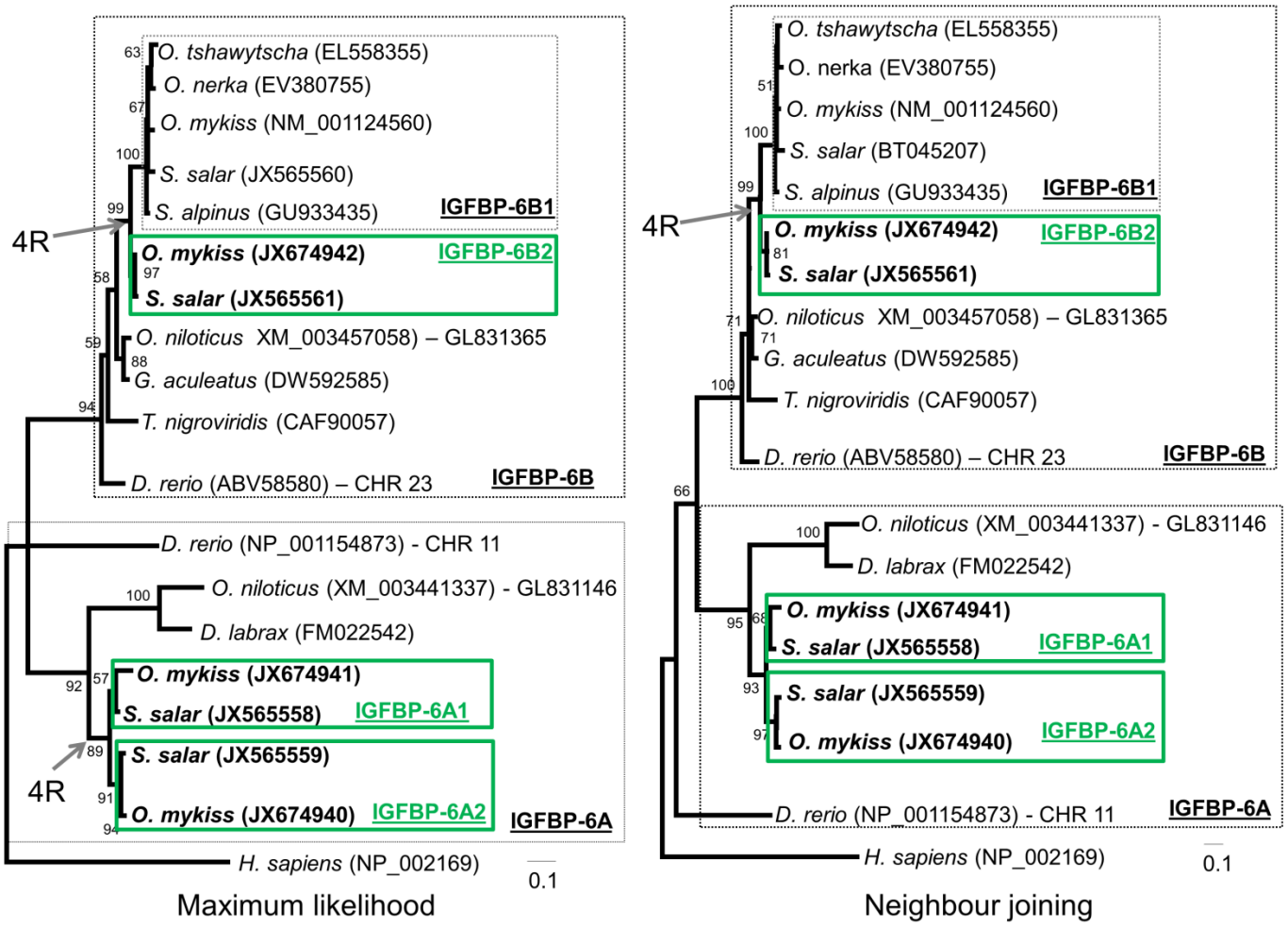

Figure S5. IGFBP-6 family member ML and NJ trees. Neither tree meets the criteria set out in the main text for 3R, although both recapture expectations of 4R and contain one teleost-wide clade matching expected systematics. In both trees, a second teleost clade contains Protacanthoptergii and Acanthopterygii sister groups, whereas a second Ostariophysi sequence branches variably. By ML this sequence is the basal branch of a monophyletic group containing Protacanthoptergii and Acanthopterygii (i.e. meeting the 3R criteria with little statistical support). By NJ, this sequence branches basally to all teleost sequences. Synteny data meets the 3R criteria set out in the main text (Fig. S6), leading us to believe that there was insufficient phylogenetic signal to recapture the correct topology. In this respect, IGFBP-6 is the shortest family member protein and had the smallest number of conserved sites in the alignment. Thus, we suggest that IGFBP-6 duplicated during 3R to produce IGFBP-6A and 6B (after Wang et al. 2009), before each paralogue-pair duplicated again during 4R producing IGFBP-6A1, -6A2, -6B1 and -6B2.

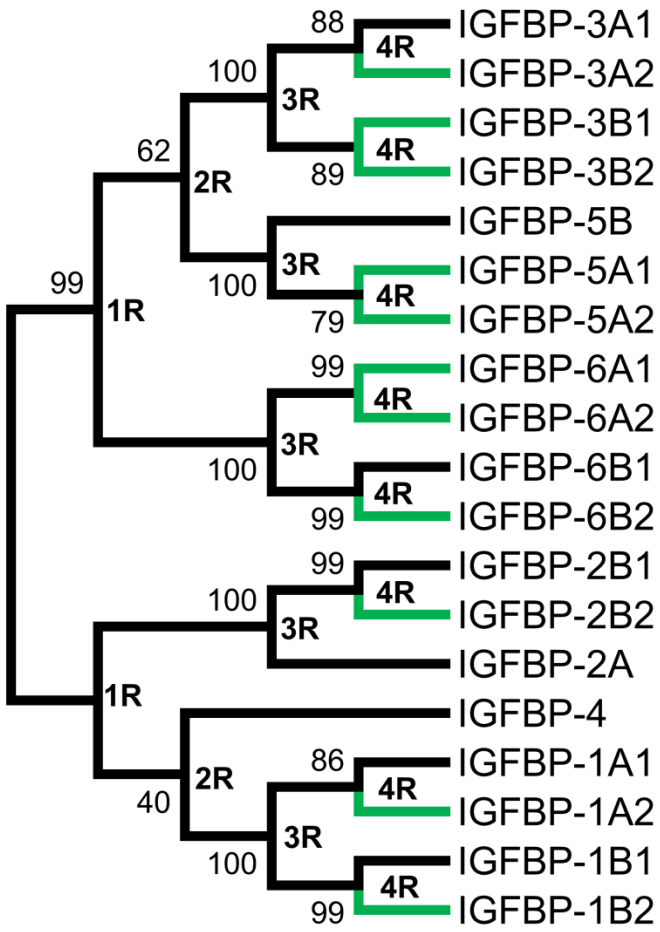

Figure S6. NJ tree produced from the same data as the ML tree provided/described in the main text, Fig. 1B. Green branches show genes discovered during this study.

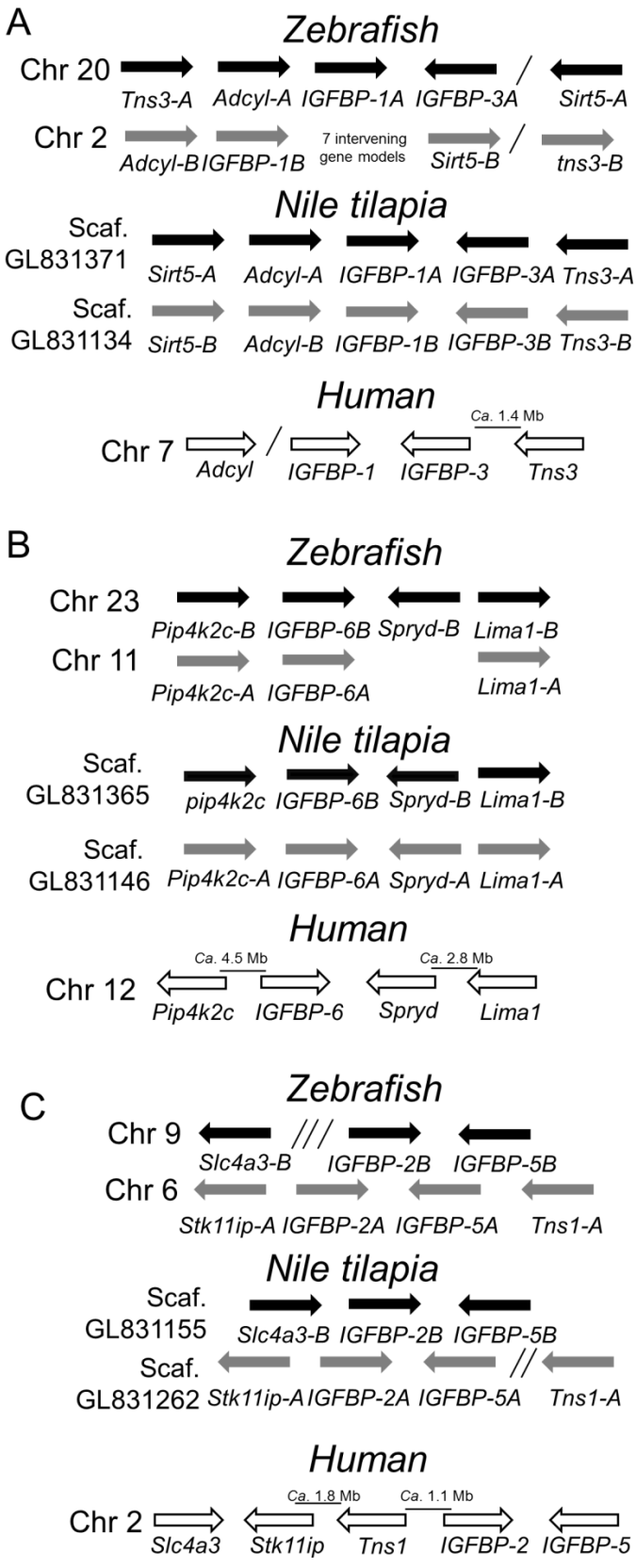
A
Zebrafish
Chr 20
Tns3-A
Adcyl-A
IGFBP-1A
IGFBP-3A
Sirt5-A
Chr 2
Adcyl-B
IGFBP-1B
7 intervening gene models
Sirt5-B
tns3-B
Nile tilapia
Scaf. GL831371
Sirt5-A
Adcyl-A
IGFBP-1A
IGFBP-3A
Tns3-A
Scaf. GL831134
Sirt5-B
Adcyl-B
IGFBP-1B
IGFBP-3B
Tns3-B
Human
Ca. 1.4 Mb
Chr 7
Adcyl
IGFBP-1
IGFBP-3
Tns3
B
Zebrafish
Chr 23
Pip4k2c-B
IGFBP-6B
Spryd-B
Lima1-B
Chr 11
Pip4k2c-A
IGFBP-6A
Lima1-A
Nile tilapia
Scaf. GL831365
pip4k2c
IGFBP-6B
Spryd-B
Lima1-B
Scaf. GL831146
Pip4k2c-A
IGFBP-6A
Spryd-A
Lima1-A
Human
Ca. 4.5 Mb
Ca. 2.8 Mb
Chr 12
Pip4k2c
IGFBP-6
Spryd
Lima1
C
Zebrafish
Chr 9
Slc4a3-B
IGFBP-2B
IGFBP-5B
Chr 6
Stk11ip-A
IGFBP-2A
IGFBP-5A
Tns1-A
Nile tilapia
Scaf. GL831155
Slc4a3-B
IGFBP-2B
IGFBP-5B
Scaf. GL831262
Stk11ip-A
IGFBP-2A
IGFBP-5A
Tns1-A
Human
Ca. 1.8 Mb
Ca. 1.1 Mb
Chr 2
Slc4a3
Stk11ip
Tns1
IGFBP-2
IGFBP-5

Figure S7. Synteny maps showing chromosomal regions containing the complete set of *IGFBP* genes for zebrafish (*D. rerio*), Nile tilapia (*O. niloticus*) and human (*H. sapiens*). Diagonal lines represent single genes that are not present at the same regions for other species. (A) Shows regions proximal to head-to-head *IGFBP-1* and *-3* genes. (B) Shows regions proximal to *IGFBP-6* genes. (C) Shows regions proximal to head-to-head *IGFBP-2* and *-5* genes. This data was used to help interpret the occurrence of 3R along with family member phylogenetic analyses (see main text Fig. 1, Figs S1-S5).

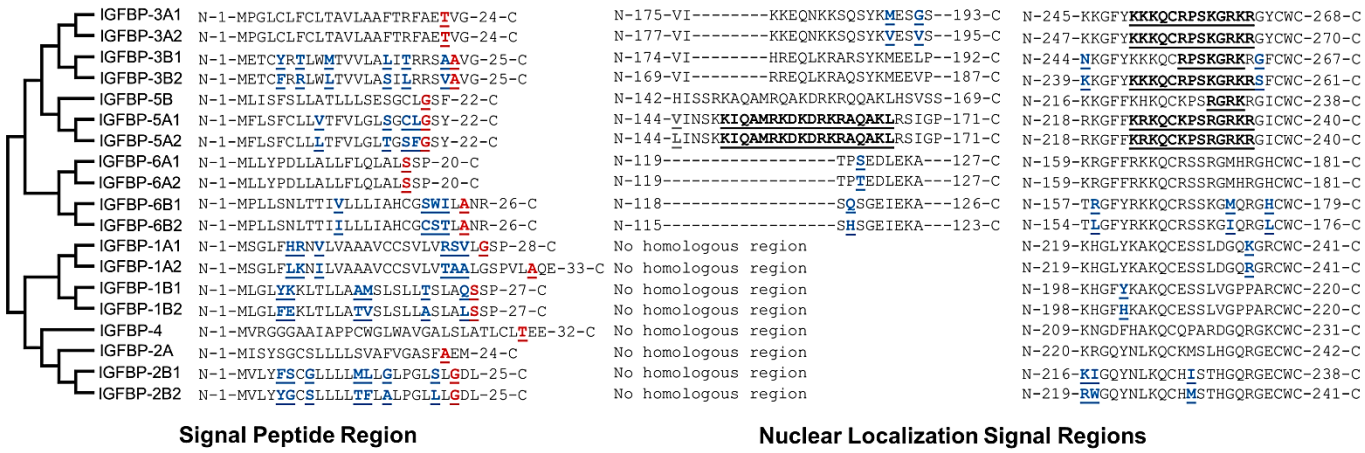

Figure S8. Predicted signal peptides and nuclear localization signals for the complete Atlantic salmon IGFBP repertoire. Proteins are mapped onto the global ML phylogenetic analysis (main text, Fig. 1B). Amino acids shaded blue have diverged between 4R paralogues, while those shaded red show predicted signal peptide cleavage sites. Broad signal peptide (SP) divergence outcomes have evolved since 4R.

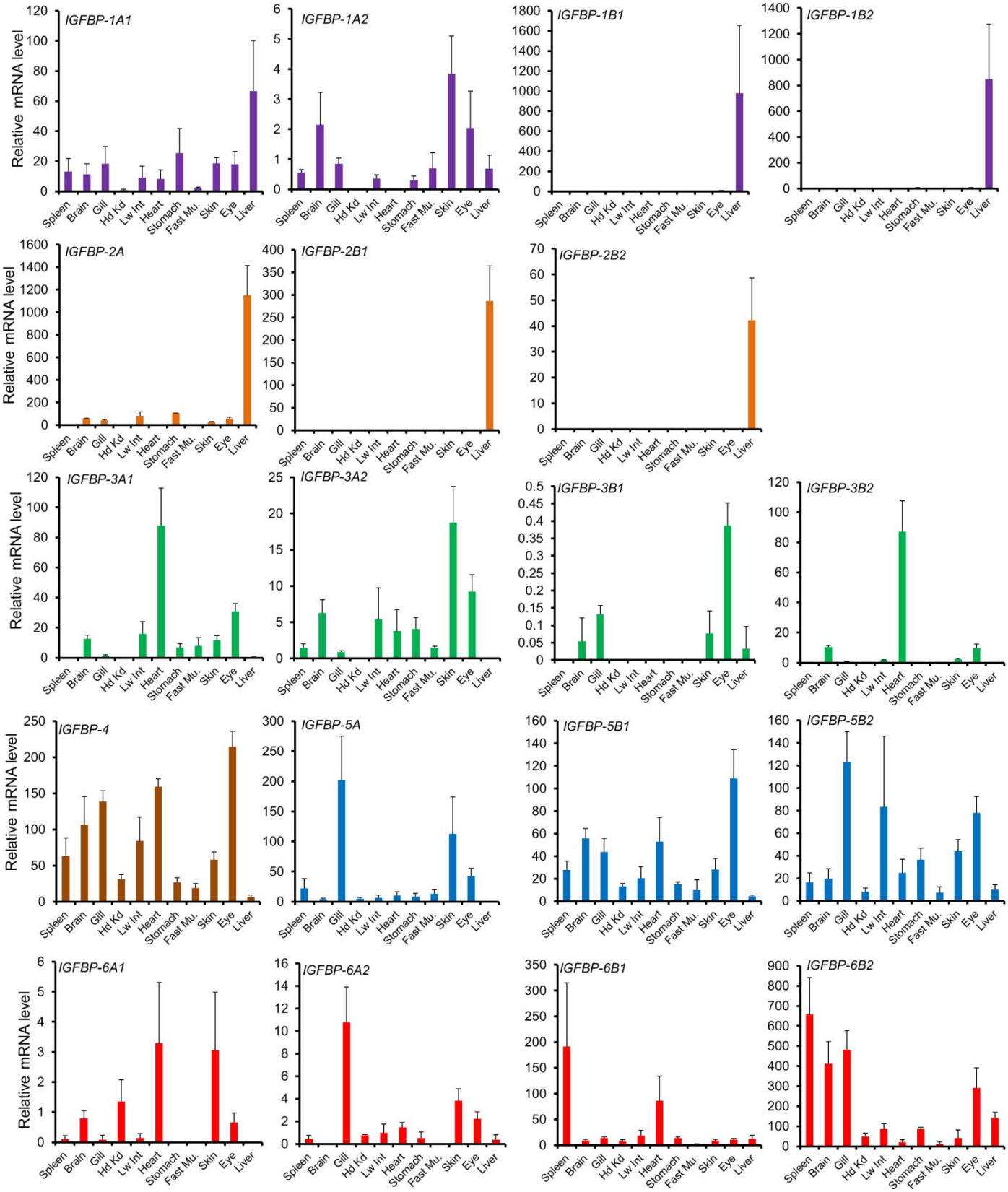
IGFBP-1A1
IGFBP-1A2
IGFBP-1B1
IGFBP-1B2
IGFBP-2A
IGFBP-2B1
IGFBP-2B2
IGFBP-3A1
IGFBP-3A2
IGFBP-3B1
IGFBP-3B2
IGFBP-4
IGFBP-5A
IGFBP-5B1
IGFBP-5B2
IGFBP-6A1
IGFBP-6A2
IGFBP-6B1
IGFBP-6B2
Relative mRNA level
Spleen
Brain
Gill
Hd Kd
Lw Int
Heart
Stomach
Fast Mu.
Skin
Eye
Liver

Figure S9. Bar charts representing the same *IGFBP* mRNA expression data shown in the main text (Fig. 4 part A) with each gene having its own chart. Error bars represent the standard deviation of the mean.

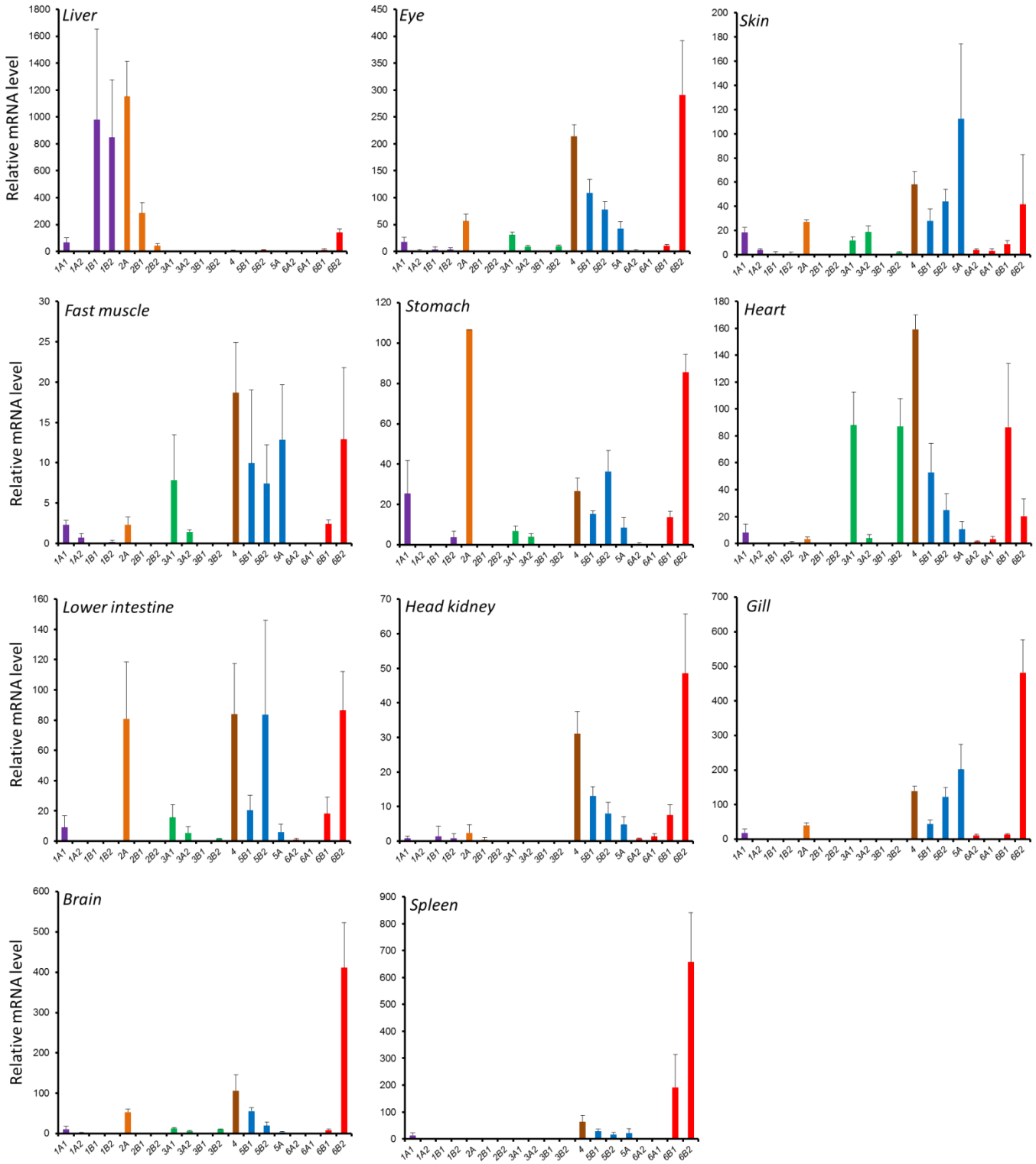

Figure S10. Bar charts representing the same *IGFBP* mRNA expression data shown in the main text (Fig. 4 part A) with each tissue having its own chart. Error bars represent the standard deviation of the mean.

References

(other than those cited in main text)

Wang X, Lu L, Li Y, Li M, Chen C, Feng Q, Zhang C, Duan C. 2009. Molecular and functional characterization of two distinct IGF binding protein-6 genes in zebrafish. Am J Physiol Regul Integr Comp Physiol. 296: R1348-R1357.