Evolution of Ancient Functions in the Vertebrate Insulin-Like Growth Factor System Uncovered by Study of Duplicated Salmonid Fish Genomes

Daniel J. Macqueen,*[1,2] Daniel Garcia de la serrana,[1] and Ian A. Johnston[1]

[1]Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, United Kingdom

[2]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, United Kingdom

***Corresponding author:** E-mail: daniel.macqueen@abdn.ac.uk.

Associate editor: William Jeffery

Abstract

Whole-genome duplication (WGD) was experienced twice by the vertebrate ancestor (2 rounds; 2R), again by the teleost fish ancestor (3R) and most recently in certain teleost lineages (4R). Consequently, vertebrate gene families are often expanded in 3R and 4R genomes. Arguably, many types of "functional divergence" present across 2R gene families will exceed that between 3R/4R paralogs of genes comprising 2R families. Accordingly, 4R offers a form of replication of 2R. Examining whether this concept has implications for molecular evolutionary research, we studied insulin-like growth factor (IGF) binding proteins (IGFBPs), whose six 2R family members carry IGF hormones and regulate interactions between IGFs and IGF1-receptors (IGF1Rs). Using phylogenomic approaches, we resolved the complete IGFBP repertoire of 4R-derived salmonid fishes (19 genes; 13 more than human) and established evolutionary relationships/nomenclature with respect to WGDs. Traits central to IGFBP action were determined for all genes, including atomic interactions in IGFBP–IGF1/IGF2 complexes regulating IGF–IGF1R binding. Using statistical methods, we demonstrate that attributes of these protein interfaces are overwhelming a product of 2R IGFBP family membership, explain 49–68% of variation in *IGFBP* mRNA concentration in several different tissues, and strongly predict the strength and direction of *IGFBP* transcriptional regulation under differing nutritional states. The results support a model where vertebrate IGFBP family members evolved divergent structural attributes to provide distinct competition for IGFs with IGF1Rs, predisposing different functions in the regulation of IGF signaling. Evolution of gene expression then acted to ensure the appropriate physiological production of IGFBPs according to their structural specializations, leading to optimal IGF-signaling according to nutritional-status and the endocrine/local mode of action. This study demonstrates that relatively recent gene family expansion can facilitate inference of functional evolution within ancient genetic systems.

Key words: **evolutionary genomics, functional evolution, gene family expansion, genome duplication, insulin-like growth factor system, insulin-like growth factor binding proteins.**

Article

Introduction

The gene content of eukaryotic genomes is organized into families related by duplication events of varying age and scale. In many lineages, a major component of this structure results from more-or-less ancient whole-genome duplication (WGD) involving tetraploidization. Heritable ploidy doubling occurs with surprising frequency (Otto and Whitton 2000; Otto 2007) and occurred in the deep ancestry of several modern multicellular and unicellular groups, sometimes in successive rounds (R) (Van de Peer et al. 2009). After WGD, the genome experiences diploidization, involving massive loss of paralogous DNA (Wolfe 2001), but nevertheless, will subsequently retain a sizeable proportion of duplicated genes (Jaillon et al. 2004; Koop and Davidson 2008; Putnam et al. 2008). These paralogs can diverge in coding and regulatory sequences during evolution and have the potential to acquire new or specialized functions (Ohno 1970; Taylor and Rae 2004; Conant and Wolfe 2008).

The Salmonidae fish family has four WGD events in its evolutionary history, namely 1R and 2R, experienced in close succession by the chordate ancestor to jawed and possibly jawed/cyclostome vertebrates (Putnam et al. 2008; Van de Peer et al. 2009); 3R, experienced later by the common teleost ancestor (Jaillon et al. 2004); and 4R, from which all extant salmonids were derived (Allendorf and Thorgaard 1984). The 4R of salmonids is associated with 30–70% paralog retention (Koop and Davidson 2008), resulting in expansions to existing vertebrate gene families including *akirin* (Macqueen, Kristjánsson, et al. 2010) and members of the ancient *Hox* system (Mungpakdee et al. 2008).

The presence of numerous 4R paralogs makes salmonids an excellent model to investigate evolution following WGD (Davidson et al. 2010; Leong et al. 2010). We envisage that the associated gene family expansion may also help interrogate functions that evolved *before* 4R. The underlying premise is that the extent to which two phylogenetically related yet independently inherited genes differ in their functions is overwhelmingly a product of evolutionary time and the associated opportunity for evolution of sequences coding functional traits. 2R family members are separated by up to 600 My

(Putnam et al. 2008), whereas their salmonid 4R paralogs are separated by 25–100 My (Allendorf and Thorgaard 1984). Consequently, differences in protein structure (and associated traits like enzymatic-function or interactions with other proteins) present across 2R gene families should typically exceed that present between 4R paralogs of member genes from the same 2R families. On these grounds, paralog retention from 4R may provide viable statistical replication of 2R, at least for certain gene functions, opening up the use of established statistical methods to study core vertebrate gene families.

To empirically explore this concept, we chose the insulin-like growth factor (IGF) binding proteins (IGFBP) gene family because its 2R origins are demonstrated (Ocampo Daza et al. 2011) and the divergent functions of six core vertebrate family members (IGFBP-1 to -6) are a product of established molecular traits. Although all IGFBPs bind essential IGF hormones with high affinity to regulate growth (Denley et al. 2005), the specifics and regulation of this interaction is presumably determined by ancient sequence evolution. Evolved differences must set the level of competition between IGFBPs and IGF1R for IGF (Siwanowicz et al. 2005; Sitar et al. 2006) and underlie distinct posttranslational modifications and interactions with proteins regulating IGFBP proteolysis and the association of IGFBPs with cell-surfaces and extracellular matrixes (Clemmons 2001). Consequently, although all IGFBP family members are efficient carriers of circulating/interstitial IGF, they have individual roles modulating the hormones delivery and interaction with IGF1R, with "functions" ultimately ranging from inhibition to potentiation of IGF-signaling (Firth and Baxter 2002; Duan and Xu 2005). The importance of IGFBP family members in different physiological contexts also depends on whether the action of IGF-regulation is systemic or local and is intimately associated with nutritional status (Clemmons and Underwood 1991; Lemozy et al. 1994; Underwood 1996; Bower et al. 2008; Shimizu, Kishimoto, et al. 2011; Shimizu, Suzuki, et al. 2011). Although experimental tools are available to infer/compare such key molecular traits across expanded IGFBP systems of salmonids, only eight genes are currently recognized in this lineage (Kamangar et al. 2006; Bower et al. 2008; Shimizu, Kishimoto, et al. 2011; Shimizu, Suzuki, et al. 2011), when many more are expected in light of 4R.

Our first objective was therefore to identify and experimentally validate the complete *IGFBP* gene repertoire of Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) before resolving evolutionary relationships with respect to known WGDs. The second objective was to generate/characterize for all identified genes, high-quality structural models of IGFBP–IGF complexes, and exhaustive quantitative mRNA expression data under physiological contexts with different requirements for IGF signaling. The final and main objective was to use statistical approaches—made possible by 4R gene expansion—to investigate whether the evolution of divergent functions among core vertebrate IGFBP family members can be uncovered by study of expanded gene repertories within salmonid genomes.

Results

Expanded Salmonid *IGFBP* Repertoires

Exhaustive bioinformatic screens of publically available nuclear genome and in-house transcriptome assemblies identified full-length protein coding sequences of 19 unique Atlantic salmon *IGFBP* genes sharing no more than 93% nucleotide sequence identity in pairwise combination. Thirteen sequences absent from the National Center for Biotechnology Information (NCBI) nucleotide database were polymerase chain reaction (PCR)-amplified and sequenced (totaling 10,185 bp, accession numbers in table 1). Eleven novel salmonid genes were identified, representing putative paralogs of all core family members except *IGFBP-4*, where a single sequence was identified. Ten of the 11 novel Atlantic salmon sequences were retrieved in rainbow trout, either by experimental sequencing or mining of assembled sequence data.

Evolutionary Relationships and IGFBP Nomenclature

Maximum-likelihood (ML) and neighbor-joining (NJ) phylogenetic analyses were performed separately for IGFBP family members -1, -2, -3, -5, and -6, based on high-confidence amino acids alignments and best-fitting models of substitution (fig. 1; supplementary material S1 and figs. S1–S5, Supplementary Material online). These exhaustive reconstructions included all known salmonid IGFBPs ($n = 47$) many additional teleost IGFBPs ($n = 52$; including all the IGFBPs predicted in the genomes of distantly related 3R species) and the full IGFBP repertoire of three tetrapod lineages ($n = 14$). A "global" phylogenetic analysis of the 19 Atlantic salmon IGFBPs was concurrently performed using an alignment of conserved N- and C-terminal regions (fig. 1*B*; supplementary material S1 and fig. S6, Supplementary Material online). We also examined conserved synteny between genomic regions containing all the *IGFBP* genes of zebrafish (*Danio rerio*, Ostariophysi), Nile tilapia (*Oreochromis niloticus*, Acanthopterygii), and human (*Homo sapiens*, Tetrapoda) (supplementary material S1 and fig. S7, Supplementary Material online).

In every family member tree, tetrapods and teleosts were monophyletic, indicating that only true teleost IGFBP orthologs of human IGFBPs were included (fig 1; supplementary material S1 and figs. S1–S5, Supplementary Material online). Observed expansions in 2R gene family structure were interpreted with respect to the following criteria: 1) that 3R should be recaptured in phylogenetic trees by two statistically supported IGFBP clades represented by the included teleost taxa (Salmonidae, Ostariophysi, and Acanthopterygii), branching according to established molecular systematics (after Near et al. 2012); 2) that 3R should be recaptured by two *IGFBP* paralogs present in Ostariophysi and Acanthopterygii genomes, located on two syntenic chromosomal regions each sharing synteny with a single human region; and 3) that 4R should be recaptured in phylogenetic trees by two statistically supported IGFBP clades, represented by at least two species of the included salmonid subfamily Salmoninae (i.e., trout, salmon, and charr species).

Table 1. Details of the Complete Salmonid *IGFBP* Gene System, Including Nomenclature, Gene Structures, Primary Protein Sequences, and Identity between Paralogs of 2R Family Members.

Proposed Salmonid Name	Previous Salmonid Name/s	NCBI Accession (Atlantic Salmon)	Gene Length, Intron–Exon Structure[a]	Amino Acid Length/Molecular Weight (kDa)	Amino Acid % Identity with 3R/4R Relatives
IGFBP-1A1	*IGFBP-1A*[b]	cDNA: JX565543 gDNA: AGKD01032414	2,030 bp. E1–379, I1–291, E2–173, I2–502, E3–129, I3–426, E4–134	268/28.7	79 vs. 1A2 57 vs. 1B1 57 vs. 1B2
IGFBP-1A2	Novel	cDNA: JX565544 gDNA: AGKD01161668	>1,835 bp. E1–379, I1–254, E2–173, I2–399, E3–129, I3–>501, E4–126	268/28.7	79 vs. 1A1 54 vs. 1B1 53 vs. 1B2
IGFBP-1B1	*IGFBP-1*[c,d] *IGFBP-1B*[b]	cDNA: JX565545 gDNA: AGKD01000719	6,149 bp. E1–397, I1–172, E2–140, I2–5021, E3–129, I3–170, E4–120	245/26.5	85 vs. 1B2 57 vs. 1A1 54 vs. 1A2
IGFBP-1B2	Novel	cDNA: JX565546 gDNA: AGKD01088518	1,866 bp. E1–366, I1–343, E2–140, I2–552, E3–129, I3–227, E4–126	247/26.5	85 vs. 1B1 57 vs. 1A1 53 vs. 1A2
IGFBP-2A	IGFBP-2[c,d]	cDNA: JX565547 gDNA: AGKD01016051 AGKD01019007	>10,353 bp. E1–365, I1–>6,555, E2–183, I2–2,248, E3–197, I3–721, E4–159	280/31.1	61 vs. 2B1 64 vs. 2B2
IGFBP2-B1	IGFBP-2B[e]	cDNA: JX565548 gDNA: AGKD01042581 AGKD01190121	>11,278 bp. E1–413, I1–>9,037, E2–147, I2–1,384, E3–141, I3–>156, E4–180	283/31.6	77 vs. 2B2 61 vs. 2A
IGFBP2-B2	Novel	cDNA:JX565549 gDNA: AGKD01106875 AGKD01074801 AGKD01075328	>29,024 bp. E1–365, I1–>3,258, E2–183, I2–8,670, E3–177, I3–1,430, E4–141	286/32.1	77 vs. 2B1 64 vs. 2A
IGFBP-3A1	IGFBP-3[e]	cDNA: JX565550 gDNA: AGKD01052954 AGKD01064800 AGKD01079623	>23,610 bp. E1–364, I1–>15,854, E2–277, I2–718, E3–120, I3–>6,151, E4–126	296/31.8	81 vs. 3A2 59 vs. 3B1 62 vs. 3B2
IGFBP-3A2	Novel	cDNA: JX565551 gDNA: AGKD01110782 AGKD01363123 AGKD01394173 AGKD01187220	>13,204 bp. E1–364, I1–>6,934, E2–284, I2–>2,196, E3–120, I3–2,468, E4–126	294/32	81 vs. 3A1 57 vs. 3B1 58 vs. 3B2
IGFBP-3B1	Novel	cDNA: JX565552 gDNA: AGKD01000719	8,480 bp. E1–370, I1–1,654, E2–266, I2–3,337, E3–120, I3–2,606, E4–127	93/32.4	83 vs. 3B2 59 vs. 3A1 57 vs. 3A2

(continued)

Table 1. Continued

Proposed Salmonid Name	Previous Salmonid Name/s	NCBI Accession (Atlantic Salmon)	Gene Length, Intron–Exon Structure[a]	Amino Acid Length/Molecular Weight (kDa)	Amino Acid % Identity with 3R/4R Relatives
IGFBP-3B2	Novel	cDNA: JX565553 gDNA: AGKD01022138	11,644 bp. E1–364, I1–1,468, E2–257, I2–1,364, E3–120, I3–7,945, E4–126	288/32.0	83 vs. 3B1 62 vs. 3A1 58 vs. 3A2
IGFBP4	IGFBP-4[c,d]	cDNA: JX565554 gDNA: AGKD01005530	12,679 bp. E1–361, I1–9,459, E2–164, I2–574, E3–126, I3–1,857, E4–138	262/28.7	N/A
IGFBP-5B1	IGFBP-5[c]	cDNA: JX565555 gDNA: AGKD01005278	13,537 bp. E1–334, I1–12,362, E2–224, I2–181, E3–120, I3–178, E4–138	271/30.0	96 vs. 5B2 74 vs. 5A
IGFBP-5B2	IGFBP-5.1[d]	cDNA: JX565556 gDNA: AGKD01022725 AGKD01053266	>10,612 bp. E1–334, I1–>9475, E2–227, I2–182, E3–120, I3–142, E4–132	270/29.9	96 vs. 5B1 72 vs. 5A
IGFBP-5A	IGFBP-5.2[d]	cDNA: JX565557 gDNA: AGKD01113827 AGKD01019007	>5,800 bp. E1–334, I1–>208, E2–203, I2–152, E3–120, I3	269/29.6	74 vs. 5B1 72 vs. 5B2
IGFBP-6A1	Novel	cDNA: JX565558 gDNA: AGKD01008228	>4,784, E4–1,720 bp. E1–343, I1–723, E2–36, I2–143, E3–121, I3–250, E4–104	204/21.9	89 vs. 6A2 52 vs. 6B1 56 vs. 6B2
IGFBP-6A2	Novel	cDNA: JX565559 gDNA: AGKD01099665	1,723 bp. E1–343, I1–712, E2–34, I2–165, E3–123, I3–232, E4–114	204/22.0	89 vs. 6A1 51 vs. 6B1 55 vs. 6B2
IGFBP-6B1	IGFBP-6[c,d]	cDNA: JX565560 gDNA: AGKD01054596	2,953 bp. E1–340, I1–1076, E2–35, I2–787, E3–120, I3–424, E4–170	199/21.5	89 vs. 6B2 52 vs. 6A1 51 vs. 6A2
IGFBP-6B2	Novel	cDNA: JX565561 gDNA: AGDK01135063	3,174 bp. E1–331, I1–1053, E2–35, I2–953, E3–120, I3–568, E4–114	202/21.9	89 vs. 6B1 56 vs. 6A1 55 vs. 6A2

[a]Spans the whole coding sequence only. "E" and "I" respectively denote exon and introns, such that E1 and I1 would describe exon 1 and intron 1, respectively, on the sense-strand, with the number after the en dash (e.g., E1–379) reflecting the nucleotide length.

[b]Shimizu, Kishimoto, et al. (2011).

[c]Kamangar et al. (2006).

[d]Bower et al. (2008).

[e]Shimizu, Suzuki, et al. (2011).

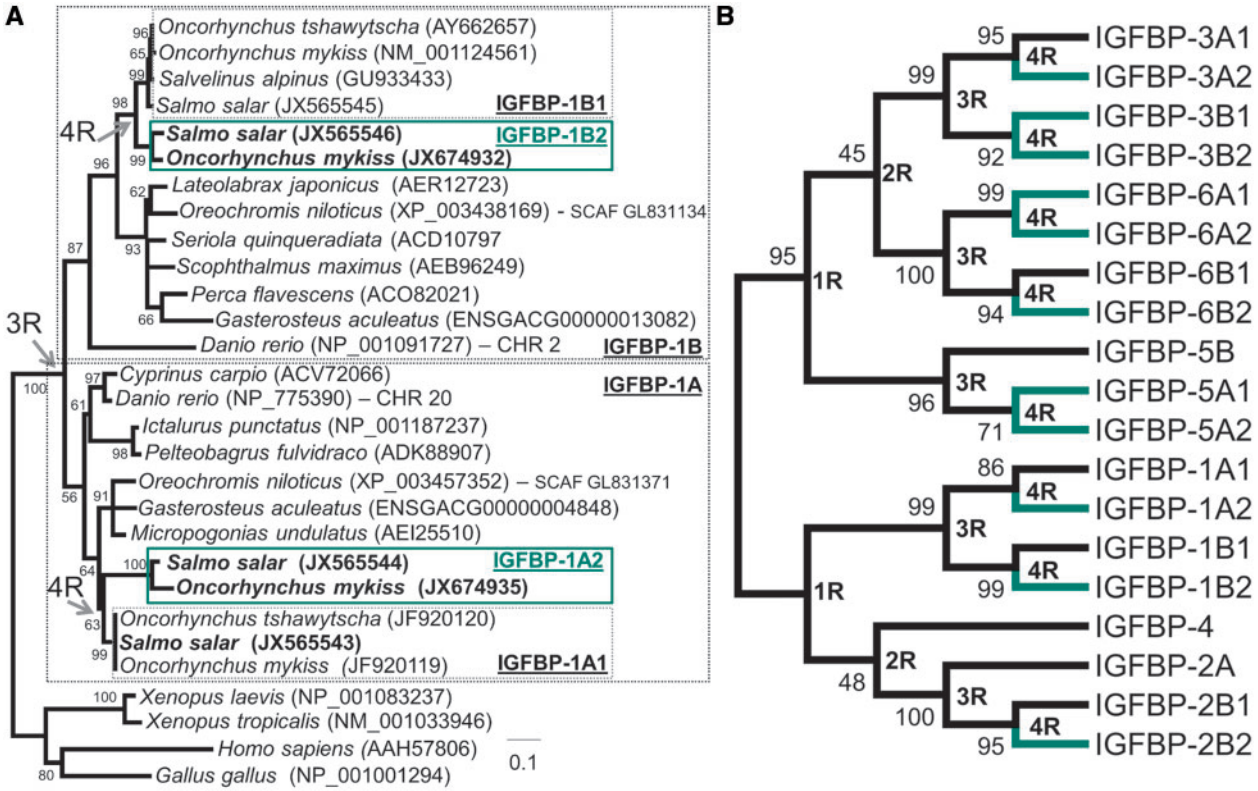

Fig. 1. (*A*) ML family member tree for IGFBP-1; ML/NJ trees for other family members are provided in the supplementary material S1 and figs. S1–S5, Supplementary Material online. The positions of 3R and 4R were inferred according to criteria set out in the main text. In this tree, the topology is consistent with duplication of *IGFBP-1* during 3R, producing *IGFBP-1A* and *-1B* (after Kamei et al. 2008) before these genes duplicated again during 4R producing *IGFBP-1A1*, *-1A2*, *-1B1*, and *-1B2* (our 4R nomenclature, table 1). Node bootstrap support values exceeding 50% are shown. Accession numbers and Ensembl gene identifiers are provided. Novel genes are boxed in green and novel sequences highlighted bold. The scale represents the number of inferred substitutions per site. (*B*) ML tree of the complete Atlantic salmon IGFBP gene system, which recaptures 3R and 4R inferred from family member reconstructions (*A*; supplementary material S1 and figs. S1–S5, Supplementary Material online). The positions of 1R and 2R are based on comparative genomics (after Ocampo Daza et al. 2011). Branching patterns within two evident IGFBP "metaclades" are sensitive to the tree-building method and statistically poorly supported (compare *B* and supplementary material S1 and fig. S7, Supplementary Material online). Green branches highlight novel salmonid *IGFBP* genes.

Consideration of the combined data led us to assign all 19 salmonid IGFBP genes to 3R and 4R (fig. 1; supplementary material S1 and figs. S1–S5, Supplementary Material online). Notably, the global analyses supported 3R/4R relationships inferred from family member trees with high statistical support (fig. 1*B*, supplementary material S1 and fig. S6, Supplementary Material online) and also indicated that IGFBP family members can be confidently separated into two "metaclades," representing a local duplication before 1R (Ocampo Daza et al. 2011) that created genes ancestral to *IGFBP-1/-2/-4* and *IGFBP-3/-5/-6*, respectively (fig. 1*B*; supplementary material S1 and fig. S7, Supplementary Material online). 3R paralogs were given the annex "A" or "B" matching existing nomenclature when orthology to relevant species was supported (supplementary material S1 and figs. S1–S5, Supplementary Material online), whereas 4R paralogs were given "1" and "2" annexes after A and B (nomenclature in table 1).

Although not a major study objective, we describe primary sequence and genomic features of salmonid IGFBPs to aid interpretation of later sections and facilitate readers wishing to further characterize these genes in the future (table 1 and supplementary material S2, Supplementary Material online).

Family-Member Characteristics of IGFBP–IGF Complexes

Homology-based structural modeling was used to infer complexes formed between the 19 Atlantic salmon IGFBPs and mature IGF1 and IGF2 hormones. In terms of incorporating the potential duplication of IGFs into the study, we initially note that there is no evidence for 3R copies of IGF1 in the literature, a notion supported by our own extensive bioinformatic screens of teleost sequence resources. Further, although duplicated putative 4R copies of IGF1 have been identified in two *Oncorhynchus* species (Wallis and Devlin 1993; Kavsan et al. 1994), we failed to identify more than a single Atlantic salmon IGF1 sequence during our own bioinformatic searches of nuclear genome and transcriptome assemblies. Importantly, the known 4R IGF1 paralogs are extremely similar in their sequences, with the mature hormones being 100% identical (Wallis and Devlin 1993). Therefore, even if an unidentified 4R IGF1 paralog does exist in Atlantic salmon, this should not affect our modeling results. We also note that while zebrafish retain duplicated 3R copies of IGF2 (Zhou et al. 2008), our own bioinformatic searches identified a single IGF2 copy in salmonids, in common with teleosts of the

Acanthoptergii superorder, suggesting that both 3R and 4R paralogs have been lost during teleost evolution.

Thus, single copies of Atlantic salmon IGF1 and IGF2 were available for modeling and shared 80% and 70% respective identity with the human IGF1 template. We used a modeling pipeline that predicts protein complexes with high accuracy (Kittichotirat et al. 2009; Macqueen, Delbridge, et al. 2010). The template was the 2.1 Å resolved ternary crystal complex of human IGF1 bound separately to the conserved N- and C-terminal regions of human IGFBP-4 termed NBP-4 and CBP-4, respectively (Sitar et al. 2006). The first five residues of NBP-4 form a "thumb" that binds IGF1 in regions including residues responsible for the interaction with IGF1R (Phe23–Tyr24–Phe25; conserved in teleost IGF1) (Siwanowicz et al. 2005; Sitar et al. 2006). To access IGF1, IGF1R must displace the NBP thumb, along with the remaining NBP, which does not prevent binding to IGF1R in its own right (Kalus et al. 1998). The NBP thumb is stabilized by interactions with CBP-4 residues, which therefore contribute to the restriction of IGF1 from IGF1R (Sitar et al. 2006). The overall affinity for IGF1 depends largely on a globular binding site between residues 39–82 in NBP-4 and is stabilized by additional contacts between IGF1/NBP-4 and CBP-4 (Sitar et al. 2006).

Unfortunately, IGFBP-6 models were necessarily excluded from further analysis owing to poor inferred local model quality in the critical NBP thumb region (fully discussed in the Materials and Methods). Important features of the crystal structure were faithfully recaptured in thirty other models inferred to have equivalent high quality to the modeling template (fig. 2A). Using UCSF Chimera (Pettersen et al. 2004), we inferred the atomic-level contacts underlying the interfaces described above (data in supplementary material S3, Excel Sheets 1–3, Supplementary Material online). Based on the study objective, we employed one-way analysis of variance (ANOVA) to test the hypothesis that statistical variation in the atomic-level contacts made at these key IGFBP–IGF interfaces is greater between than within core IGFBP family members. Considering contacts underlying the NBP thumb–IGF interface, there was great support for the hypothesis for models involving both IGF1 and IGF2 (respective F ratios for IGF1 and IGF2 models = 14.2 and 13.34, $P < 0.0001$) (fig. 2B). Post hoc comparisons showed that the NBP thumb of IGFBP-1 and -2 complexes makes significantly more contacts with IGF1 or IGF2 than IGFBP-3 and -5 complexes (fig. 2B). The NBP thumb of IGFBP-1 and -2 also makes more contacts with both IGF1 and IGF2 than equivalent IGFBP-4 complexes, although, because $n = 1$ for IGFBP-4, statistical power is limited in these comparisons. There was remarkable conservation in the number of contacts made between the NBP thumb of IGFBPs and IGF1 or IGF2. This is evident in mean core family member values (fig. 2B) and the striking correlation among 15 individual IGFBPs (Pearson's $R = 0.97$, $P < 0.0001$). A notable exception was that the NBP thumb of IGFBP-4 was predicted to make around a quarter more contacts with IGF2 than IGF1; this underlies a significant difference comparing IGFBP-4–IGF2 with IGFBP-5–IGF2 complexes (fig. 2B), despite $n = 1$ for IGFBP-4.

There was also a significant IGFBP family member effect (respective F-ratios for IGF1 and IGF2 models = 7.5 and 5.26; $P = 0.005$ and 0.015, respectively) considering stabilizing contacts made between the NBP thumb and CBP (explaining 69% and 60% of the respective variation across 15 IGFBPs for IGF1 and IGF2 models), with IGFBP-1, -2, and -5 having significantly more contacts than IGFBP-3 (fig. 2B). Again, there was strong conservation in the number of contacts made between the NBP thumb and CBP comparing IGF1 with IGF2 models (fig. 2B, Pearson's $R = 0.87$, $P < 0.0001$). Interestingly, the NBP thumb and CBP of IGFBP-4 was predicted to make around a quarter fewer contacts when bound to IGF2 as opposed to IGF1 (fig. 2B).

However, there was no family member effect considering contacts made between IGF1 or IGF2 and NBP residues outside the thumb ($P = 0.956$ and 0.914, respectively, fig. 2B). Therefore, there is a remarkable contrast in how well core IGFBP family membership explains the variation in the number of contacts made between IGF and NBP residues comprising the thumb versus otherwise (respectively 82/81% and 0/0% of the total variation for IGF1/IGF2). There was strong conservation in the number of contacts made between NBP residues outside the thumb and IGF1 or IGF2 (fig. 2B, Pearson's $R = 0.92$, $P < 0.0001$).

To aid the depiction of family member differences described earlier, three example complexes involving IGF1 and focused on the NBP thumb region are shown in figure 3. Major differences were evident in the number of contacts made between the NBP thumb and Phe23, Tyr24, and Phe25 of IGF1/IGF2 (fig. 3; supplementary material S3, Excel Sheets 1 and 2, Supplementary Material online). There were also striking family member differences in the number of contacts made between the NBP thumb and IGF1/2 involving other residues (fig. 3; supplementary material S3, Excel Sheets 1 and 2, Supplementary Material online), with IGFBP-1 and -2 complexes having at least 3-fold more such contacts than IGFBP-3, -4, and -5 complexes (fig. 3). It is also interesting to note the nature of the stabilizing interface between the NBP thumb and CBP (fig. 3). For example, despite there being a similar number of contacts in IGFBP-1, -2, and -5 complexes, the CBP surface interacting with the NBP thumb is smaller and more distal from IGF in IGFBP-5 complexes, meaning no additional contacts are made between IGF and CBP (fig. 3).

Tissue Expression of the Complete Salmonid *IGFBP* System

The mRNA levels of all 19 *IGFBP* genes were measured in 11 Atlantic salmon tissues using quantitative PCR (fig. 4A; supplementary material S1 and figs. S9 and S10, Supplementary Material online). Liver had three times more sum *IGFBP* message than any other tissue, 95% of which comprised *IGFBP-1* and -2 family member transcripts, particularly liver-specific *IGFBP-1B1*, *IGFBP-1B2*, and *IGFBP-2B1* and more widely expressed *IGFBP-2A* (fig. 4A; supplementary material S1 and fig. S10, Supplementary Material online). Like its 4R paralog, *IGFBP-2B2* was liver specific, but contributed <1% to the total liver message (fig. 4A; supplementary material S1 and fig. S10,

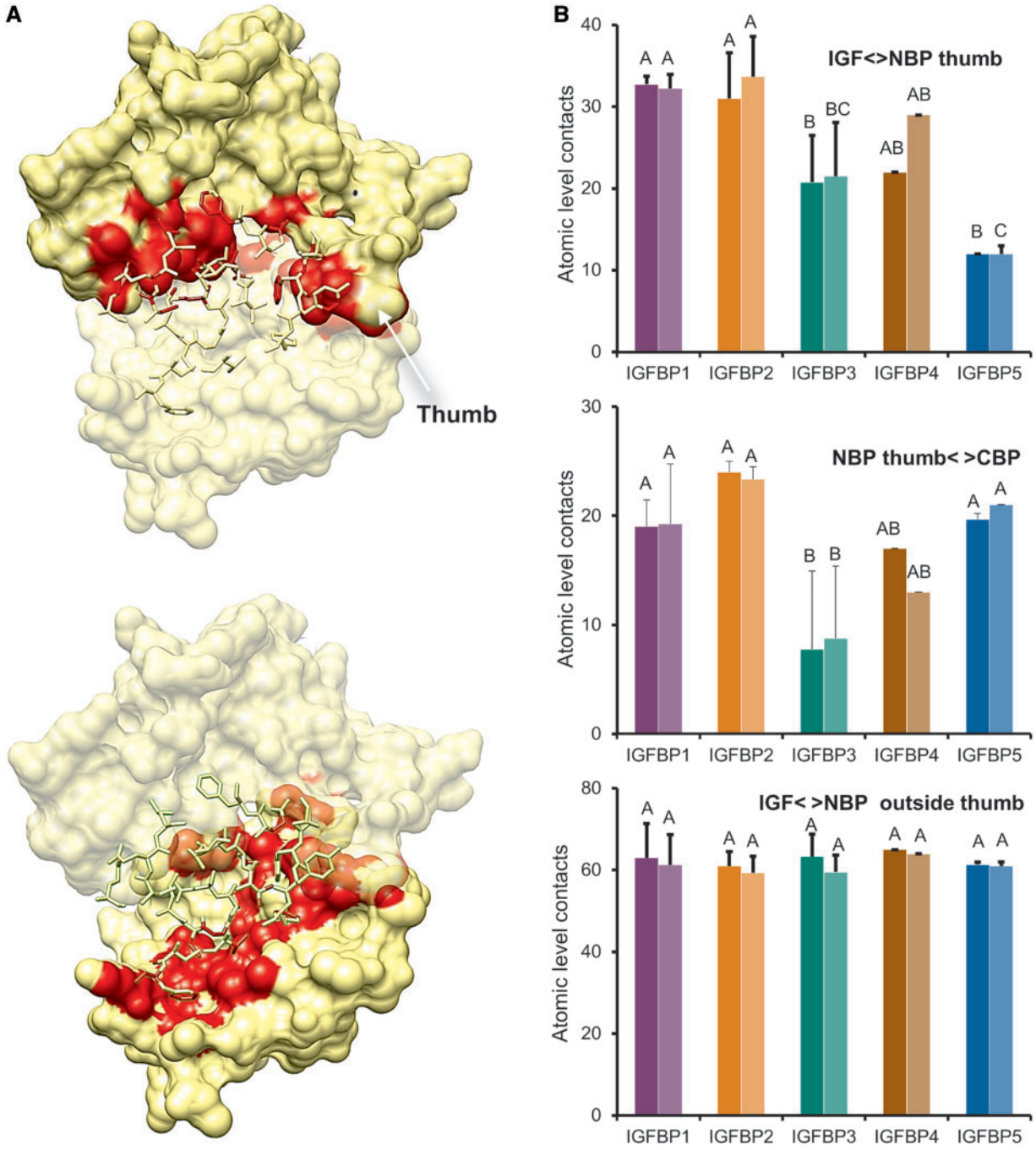

FIG. 2. (*A*) Chimera renderings of a modeled ternary complex containing Atlantic salmon IGF1, NBP-4, and CBP-4. CBP and NBP surfaces are shown (transparent in the upper and lower images, respectively). IGF1 is shown as a ribbon with residues contacting NBP or CBP having sidechains. Inferred atomic-level interactions are shaded red, between NBP–IGF1 and NBP–CBP in the upper image and CBP–IGF1 and CBP-NBP in the lower image. The NBP thumb is highlighted by an arrow and the main IGF-binding region is evident as a large patch of red shading on the NBP surface. (*B*) Bar charts comparing core IGFBP family members in terms of the number of atomic-level contacts comprising three interfaces (identified) in IGFBP–IGF1 and IGFBP–IGF2 complexes. For each IGFBP family member, the left and right hand bars show contacts made in IGFBP–IGF1 and IGFBP–IGF2 complexes. All data are means + SD with *n* equal to the 4R gene number. Different letters indicate significant differences ($P < 0.01$) between IGFBP family members compared separately for models containing IGF1 and IGF2.

Supplementary Material online). *IGFBP-1A1* was unique among *IGFBP-1* family members in being notably expressed outside liver, whereas its 4R paralog *IGFBP-1A2* was lowly expressed in all tissues (fig. 4*A*; supplementary material S1 and figs. S9 and S10, Supplementary Material online). *IGFBP-4* was more highly expressed than all other IGFBP-1/-2/-4 metaclade genes combined in many tissues, but with the exception of *IGFBP-1A2*, was 87-fold less abundant on average in liver (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online).

Among 11 genes from the IGFBP-3/-5/-6 metaclade, only *IGFBP-6B2* was expressed to any relative extent in liver,

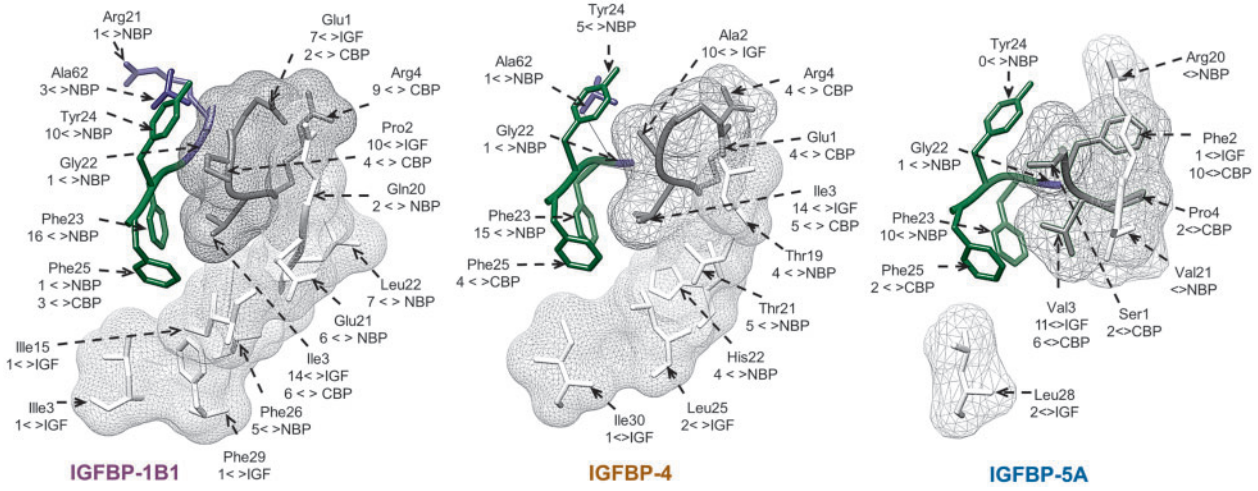

FIG. 3. Examples of the NBP thumb region of Atlantic salmon IGFBP–IGF1 complexes representing three IGFBP family members. Residues comprising the NBP thumb and CBP interface are colored gray and white, respectively, with the corresponding surfaces portrayed as meshes. IGF1 residues predicted to interact with the IGF1R are colored green (or blue otherwise). All residues are labeled with details provided of interactions with other surfaces in the complex.

comprising approximately 4% of the total *IGFBP* message (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online). This gene also accounted for 40–65% of the *IGFBP* message in spleen, brain, gill, and head-kidney (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online). In heart, *IGFBP-3A1*, *-3B2*, and *-6B1* were relatively abundant (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online). Outside heart, *IGFBP-3* genes were generally relatively lowly expressed, although *IGFBP-3A1* comprised approximately 10% the total fast-muscle *IGFBP* message (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online). Fast muscle expressed less *IGFBP* than other tissues, with 70% of the message coming from the IGFBP-3/-5/-6 metaclade (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online). *IGFBP-5* genes were also relatively abundant in skin, gill, and eye (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online). *IGFBP-3B1* and the 4R *IGFBP-6A* paralogs were expressed at relatively negligible levels in all tissues studied (fig. 4*A*; supplementary material S1 and fig. S10, Supplementary Material online).

Unsupervised hierarchical clustering analysis was used to group genes according to correlation in expression across tissues (fig. 4*B*) using data ranks (i.e., Spearman's correlation; corresponding Rho [ρ] values for 161 gene-pairs provided in supplementary material S3, Excel Sheet 4, Supplementary Material online). *IGFBP6-B2* paralogs clustered together outside all other genes, which split into two further clusters, the first comprising *IGFBP-1* and *-2* genes (fig. 4*B*) and the other containing genes from the IGFBP-3/5/6 metaclade, along with *IGFBP-4* and *IGFBP-1A2*.

Association between IGFBP–IGF-Binding Characteristics and *IGFBP* Tissue Expression

The results indicated that two interfaces in IGFBP–IGF complexes that specifically regulate IGF1–IGF1R binding have most relative contacts when *IGFBP-1* and *-2* family members

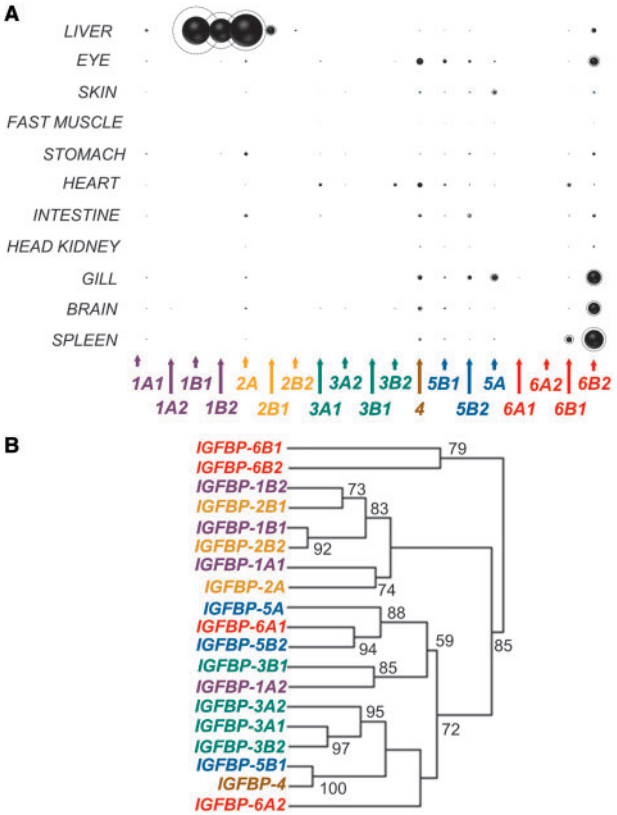

FIG. 4. mRNA expression of 19 *IGFBP* genes in 11 juvenile Atlantic salmon tissues. (*A*) qPCR-derived expression levels portrayed in the style of a northern-dot blot, scaled to be relative across genes. The area of black circles represents the mean expression level and the distance between the circumference of black and dotted circles the SD ($n = 4$). Gene-by-gene bar graphs showing the same data are provided in the supplementary material S1 and figs. S9 and S10, Supplementary Material online. Unsupervised hierarchical clustering of *IGFBP* gene expression. Numbers at branch nodes are AUP values (Suzuki and Shimodaira 2006) based on 5,000 bootstrap iterations.

are involved, whereas the same genes comprise most of the liver *IGFBP* message. To formally investigate this apparent association within a statistical framework incorporating all genes and tissues, we used regression modeling to assess whether *IGFBP* mRNA levels were predicted by the number of atomic-level contacts made between the NBP thumb and IGF (hereafter, α contacts) and between the NBP thumb and CBP (hereafter, β contacts).

In IGFBP–IGF1 complexes, α and β contacts were statistically important predictors of *IGFBP* expression from liver, fast muscle, and head-kidney, whereas solely α contacts were important predictors of expression in seven other tissues (table 2). The results were similar for IGFBP–IGF2 complexes, although the relationships were slightly weaker and β contacts had less importance for tissues outside liver (table 2). For both IGFBP–IGF1 and IGFBP–IGF2 complexes, the single strongest regression model was attained for liver expression; >63% of the expression level variation across 15 *IGFBP* genes was explained by the combined number of α and β contacts (table 2). After employing Bonferroni correction to avoid type I errors associated with the multiple comparisons, several tissues still had significant regression models, including (in addition to liver), gill, skin, eye, and fast muscle for IGFBP–IGF1 comparisons and gill and skin for IGFBP–IGF2 comparisons (table 2). These patterns were evident in scatterplots, where the variation stratified largely with core family members (fig. 5). The association between mRNA level and α/β contacts was positive for liver and negative for other tissues (table 2 and fig. 5).

Can IGFBP–IGF-Binding Characteristics Predict IGFBP Regulation When Requirements for IGF-Signaling Are Radically Altered?

We hypothesized that there may also exist an association between the number of α and β contacts and *IGFBP* transcriptional regulation according to the nutritional-status of liver, the primary source of endocrine *IGFBP* and IGF. Specifically, we rationalized that in scenarios where it is favorable to minimize investment of energetic resources into growth, *IGFBP* genes coding the most α and β contacts should be upregulated to increase competition with IGF1R for IGF1, but when active growth is favorable, the same genes will show downregulation. On similar grounds, we hypothesized that those IGFBPs having fewer relative α and β contacts are better candidates to potentiate IGF1R signaling and will therefore show a direct reciprocal pattern of transcriptional regulation. To test these a priori hypotheses, we subjected Atlantic salmon juveniles to a period of 72 h fasting, followed by 18 h feeding and measured the liver expression profiles of 10 *IGFBP* genes expressed at quantifiable levels. This short fasting period was selected to ensure that the digestive system was empty, but to avoid the extensive catabolism expected with longer periods of food restriction (Johnston, Bower, et al. 2011).

There were only minor differences in *IGFBP* expression between the livers of fish fed at nonsatiating levels or fasted for short periods (fig. 6), with no correlation observed between the number of α and β contacts and the associated mRNA-fold regulation ($P = 0.27$ and 0.55 for IGF1 and IGF2).

Table 2. Regression-Associating *IGFBP* Gene Expression Levels with the Number of Atomic-Level Contacts at Key Interfaces in IGFBP-IGF1/IGF2 Complexes.

mRNA Level	Best Regression Model	R^2 (%)	*P*	Mallow's Cp, S
IGF1				
Liver	$= -4.03 + 0.216\alpha + 0.399\beta$	67.7	0.001	3.0, 2.74
Gill	$=17.0 - 0.371\alpha$	53.9	0.002	2.3, 3.15
Skin	$=16.9 - 0.368\alpha$	52.8	0.002	2.9, 4.32
Eye	$=16.6 - 0.355\alpha$	49.2	0.004	3.9, 3.31
Fast muscle	$=12.7 - 0.373\alpha + 0.259\beta$	60.8	0.004	3.0, 3.01
Head Kidney	$=7.34 - 0.255\alpha + 0.404\beta$	55.4	0.008	3.0, 3.21
Heart	$=16.0 - 0.327\alpha$	42.0	0.009	1.4, 3.53
Lower intestine	$=15.9 - 0.323\alpha$	41.4	0.010	3.9, 3.52
Brain	$=15.8 - 0.319\alpha$	40.1	0.011	1.3, 4.40
Spleen	$=14.8 - 0.280\alpha$	31.1	0.031	1.4, 3.82
IGF2				
Liver	$= -4.43 + 0.249\alpha + 0.355\beta$	63.2	0.002	3.0, 2.93
Gill	$=16.7 - 0.341\alpha$	49.8	0.003	1.4, 3.29
Skin	$=16.6 - 0.339\alpha$	49.1	0.004	1.6, 3.31
Eye	$=15.9 - 0.310\alpha$	41.0	0.010	1.9, 3.57
Fast muscle	$=15.8 - 0.30\alpha$	40.8	0.010	3.1, 3.56
Heart	$=15.2 - 0.285\alpha$	34.7	0.021	1.0, 3.74
Lower intestine	$=15.0 - 0.275\alpha$	32.8	0.026	2.2, 3.77
Brain	$=14.9 - 0.272\alpha$	31.8	0.028	1.0, 3.82
Spleen	$=14.5 - 0.254\alpha$	28.1	0.042	2.2, 3.90

Note.—< > = Number of predicted atomic level contacts at protein interface. α = IGF1 < > NBP thumb. β = NBP thumb < > CBP. Underlined probability values remain significant after Bonferroni correction.

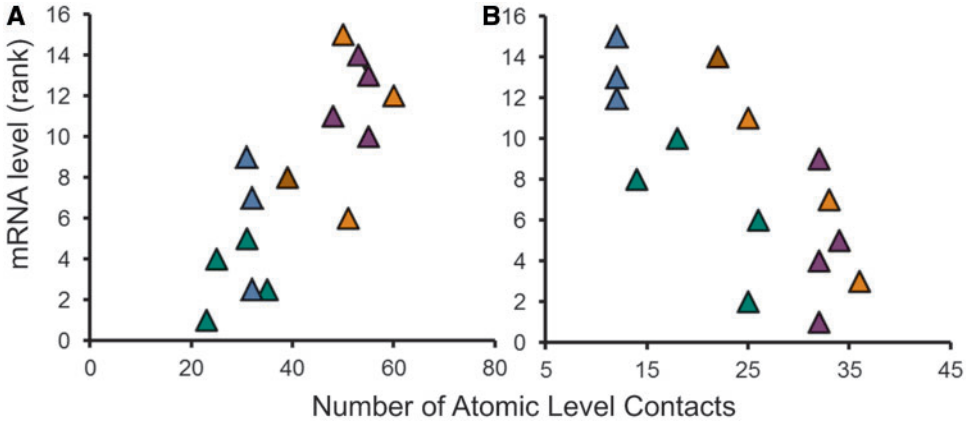

FIG. 5. Example scatterplots showing the association between atomic-level contacts within modeled IGFBP–IGF1 complexes and *IGFBP* mRNA expression in tissues. Family members are colored as in figure 2. (*A*) Illustrates how *IGFBP* liver mRNA level is positively correlated with the combined number of contacts made between the NBP thumb and IGF1 and between the NBP thumb and CBP. (*B*) Illustrates how *IGFBP* skin mRNA level is negatively associated with the number of predicted contacts between the NBP thumb and IGF1. Associated results are given in table 2.

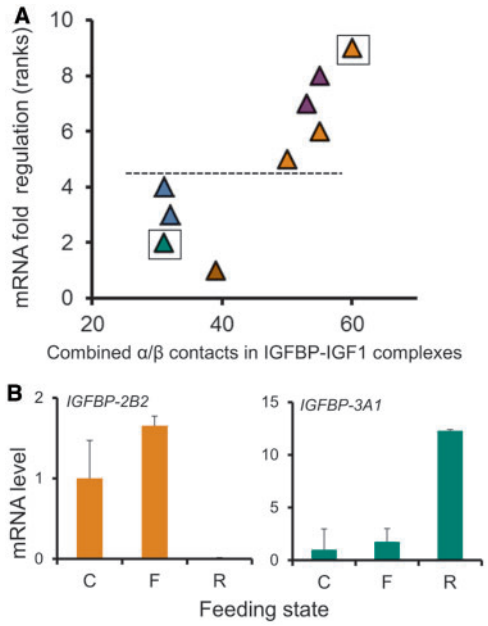

FIG. 6. (*A*) Scatterplot showing the association between the number of α and β contacts in modeled IGFBP–IGF1 ternary complexes and *IGFBP* mRNA regulation in liver at different nutritional states. The highest ranked *IGFBPs* on the *y* axis showed the greatest downregulation during 18 h ad libitum feeding that followed a 72 h period of feed restriction. *IGFBPs* above and below the dotted line were downregulated and upregulated, respectively. (*B*) Expression data for the two boxed *IGFBP* genes in (*A*). On the *x* axis, C, F, and R indicate control, fasted and refeeding states. On the *y* axis, mRNA expression level are scaled such that the C-state mean equals one; the two charts are not on equivalent scales and should only be compared to indicate the strength and direction of transcriptional regulation. Error bars represent standard deviation. IGFBP family members are colored as in figure 2.

However, our predictions were strongly supported in terms of the strength/direction of *IGFBP* family member expression observed when fasted individuals were refed to satiation (shown for IGF1 models in fig. 6*A*). The correlation between the combined number of α and β contacts and *IGFBP*-fold regulation between fasting and refeeding was significant ($\rho = 0.86$ and 0.75, $P = 0.003$ and 0.019 considering IGF1 and IGF2 models, respectively, fig. 6*A*). Considering the number of α and β contacts separately, both correlations were still significant for IGF1 models ($\rho = 0.82$ and 0.73 for α and β contacts, respectively; $P = 0.007$ and 0.025), whereas solely α contact correlations were significant for IGF2 models ($\rho = 0.69$ and 0.36 for α and β contacts, respectively; $P = 0.039$ and 0.36).

The level of *IGFBP* regulation in the predicted direction was also striking, with *IGFBP-1* and *-2* genes showing between 20-fold to 3,800-fold downregulation upon postfast refeeding and *IGFBP-3*, *-4*, and *-5* genes showing up to 30-fold upregulation during the same period (examples in fig. 6*B*).

Discussion

The NBP thumb and its contacts with IGF and CBP are fundamental structural determinants of IGF-signaling because these regions must be displaced by the IGF1R to access IGF (Sitar et al. 2006). Our modeling results suggest that the numbers of contacts comprising these interfaces are different among core vertebrate IGFBP family members, but similar for the same IGFBP complexes containing either IGF1 or IGF2. This is compatible with a scenario where sequence evolution following 1R/2R led to IGFBP family members providing distinct levels of competition with IGF1R for IGF hormones and that subsequent evolution after 3R/4R is yet to approach these boundaries of divergence. The local duplication of a proto-IGFBP is thought to have occurred after the split of urochordates and chordates, creating a tandem gene pair, that went on to separately generate *IGFBP-1/-2/-4* and *IGFBP-3/-5/-6* during 2R (Ocampo Daza et al. 2011). Although insufficient phylogenetic signal exists to infer the precise evolutionary relationships within each metaclade (compare fig. 1*B* vs. supplementary material S1 and fig. S6, Supplementary Material online) (see also Ocampo Daza et al. 2011), the conserved high and low number of α contacts, respectively, associated with salmon IGFBP-1/-2 (plus IGFBP-4 when in complex with IGF2) and IGFBP-3/-5 proteins suggest these traits could predate 2R if representative of the metaclade

ancestral states. Notably, such findings could not have been made with a 2R species, which retain single copy genes of four to six IGFBP family members (Ocampo Daza et al. 2011) because there is no statistical power. On similar lines, had we tested associations between *IGFBP* expression and IGFBP–IGF complex attributes in 2R species, even strong effects would probably remain equivocal, because the minimum R^2 value for $P = 0.05$ is 0.9, 0.77, and 0.67 when $n = 4$, 5, and 6, equivalent to a correlation statistic of 0.95, 0.88, and 0.82, respectively. If correction for multiple comparisons was required, a significant *P* value would require correlations unlikely to exist in biological systems.

Under the presented model of structural evolution, we suggest that 2R IGFBP family members were predisposed to functions with different requirements for IGF-signaling via IGF1R. Having IGFBPs specialized to distinct biological contexts would hypothetically have been selectively advantageous in the ancestral vertebrate, facilitating the evolution of increasingly complex regulation of IGF-dependent growth. The statistical association between α (and to a lesser extent β) contacts and *IGFBP* expression may therefore reflect the need to ensure the appropriate IGFBPs are produced under different physiological settings. In terms of the strong correlation observed in liver, it is notable that the IGF-bound IGFBP population in teleost circulation is predominantly a function of mRNA expression from this tissue (see supporting references and discussion in supplementary material S2, Supplementary Material online). Briefly, this statement is derived from two facts: 1) that the plasma of species spanning the teleost phylogeny contains only IGFBP-1 and IGFBP-2 family member proteins detected in complex with IGF, and 2) that in diverse teleost species, *IGFBP-1* and *-2* genes generate most of the liver *IGFBP* message and are either liver-specific or most abundantly expressed from the liver (our results; supporting supplementary material S2, Supplementary Material online). Thus, the majority of circulating IGF is carried by IGFBP family members with the greatest relative number of α/β contacts in salmon. As liver is also the predominant source of endocrine IGF, this may reduce the chance of the secreted hormone binding to liver IGF1Rs before reaching more distal target tissues. Importantly, this situation is not easily comparable with the endocrine IGFBP phenotype of mammals, where most of the circulating IGF is bound to IGFBP-3 (and to a lesser extent, IGFBP-5) as part of a larger complex containing the acid labile subunit (ALS) protein (reviewed by Boisclair et al. 2001). The size of the ALS complex physically restricts IGF to the vascular compartment and acts to increase the half-like of circulating IGF (Boisclair et al. 2001). There is no evidence for the ALS complex in species spanning the teleost phylogeny (Shimizu et al. 1999; Degger et al. 2000). Thus, in contrast to the teleost state, where circulating IGFBP–IGF complexes can freely acquire proximity to IGF1Rs, the presence of the ALS complex in mammals disconnects the direct link between liver *IGFBP* expression, the circulating IGFBP population and IGF1R signaling. Thus, even though circulating IGFBP-3 of mammals arises predominantly from liver expression (Phillips et al. 1998), the endocrine IGF in complex with IGFBP-3 cannot access cell-membrane IGF1Rs. Therefore, there is no reason to expect that IGFBP-3 structural properties (related to IGF1R signaling) should be associated with liver expression. Under parsimony, the simpler endocrine phenotype of the teleost IGF system is more akin to that of the lobe-/ray-finned fish ancestor. Teleosts are considerably more ancient than mammals, whereas lampreys also lack the ALS association (Upton et al. 1993) and predate all jawed vertebrates.

In several tissues other than the liver, an inverse correlation was observed between *IGFBP* expression level and the number of α contacts. For example, in well-fed salmon, genes from the IGFBP-3/-5/-6 metaclade and *IGFBP-4* were more highly expressed than *IGFBP-1* and *-2* in gill, skin, heart, and fast muscle, but had meagre liver expression, suggesting minor roles in circulation with predominant local actions. These predominantly local-acting IGFBPs may generally provide reduced competition with IGF1R for IGF than the systemic IGFBP population, reflecting their normal roles in the potentiation of IGF-signaling. Under this model, having a relatively lower number of α contacts would facilitate the release of the hormone to IGF1R, although this process is likely highly complex and dependent on factors not considered here (e.g., interactions between IGFBPs and cell-membrane proteins acting to concentrate IGFBP–IGF complexes near IGF1Rs). In support of this model, numerous studies have demonstrated that locally acting IGFBP-3 and -5 can augment IGF-signaling in diverse tissues and physiological contexts (Andress and Birnbaum 1992; Conover 1992; Ramagnolo et al. 1994; Ewton et al. 1998; Firth and Baxter 2002; Kiepe et al. 2002, 2006; Ren et al. 2008), whereas IGFBP-1 and -2 generally inhibit the growth-promoting functions of IGF in both mammals and fish (Gockerman et al. 1995; Lee et al. 1997; Duan et al. 1999; Firth and Baxter 2002; Kiepe et al. 2002; Kajimura et al. 2005; Zhou et al. 2008; Kamei et al. 2008). In contrast to our model, mammalian data have universally indicated that IGFBP-4 inhibits IGF-dependent IGF1R signaling (Jones and Clemmons 1995; Duan and Clemmons 1998; Ewton et al. 1998). However, there is a growing body of indirect evidence that IGFBP-4 has growth-promoting roles in salmonids (Bower et al. 2008; Bower and Johnston 2010; Macqueen et al. 2011) and other teleosts (Garcia de la serrana et al. 2012), consistent with a function promoting IGF-signaling and suggesting divergence in function during vertebrate evolution. Another interesting potential explanation for the results is that IGF-independent functions, characterized extensively for IGFBP-3 and -5 family members (reviewed by Schneider et al. 2002; Yamada and Lee 2009), but also known for IGFBP-4 (Zhu et al. 2008), have impacted on the evolution of the structural interaction between the NBP thumb and IGFs through pleiotropic mechanisms.

We increased the biological relevance of our data by verifying an a priori hypothesis that directionally predictable associations should exist between the number of α or β contacts and the regulation of liver *IGFBP* transcript-levels according to nutritional status. Indeed, the expression of *IGFBPs* with relatively fewest α/β contacts (i.e., *IGFBP-3 -4*, and *-5* genes) increased on postfast feeding, whereas *IGFBPs* with the highest number of α/β contacts (i.e., *IGFBP-1* and *-2*

genes) showed massive downregulation, to the point of nonexpression in specific cases (fig. 6). The teleost liver is highly metabolically active and functions as an initial energy store mobilized to other tissues during fasting and these resources are rapidly restored upon refeeding (Power et al. 2000). The dramatic shift in IGFBP structural attributes associated with the observed changes in liver *IGFBP* expression may act to relax competition for IGF1 with liver IGF1Rs augmenting IGF1-signaling and protein-synthesis required during metabolic recovery. Increased liver expression of *IGFBPs* with a lower number of α/β contacts should also be reflected in the circulating IGFBP population, promoting IGF-signaling systemically during the rapid compensatory growth that follows postfast feeding (Johnston et al. 2011). Functionally comparable regulation of *IGFBP-1* and *-2* genes with nutritional status has also been observed in mammals (Tseng et al. 1992; Lemozy et al. 1994; Underwood 1996) and other teleosts (Duan et al. 1999; Shimizu et al. 2006; Kamei et al. 2008), which might reflect ancestral conservation of structural properties disposed toward inhibition of IGF-signaling.

The results also have importance for the systematics of teleost *IGFBP* families. Although several published phylogenetic analyses have concluded that IGFBPs duplicated during 3R, just one considered the complete core family, concluding that 3R paralogs have been retained for IGFBP-1, -2, -3, -5, and -6 (Ocampo Daza et al. 2011). However, many presented ML/NJ branching patterns did not recapture 3R according to our criteria (see figs. within Ocampo Daza et al. 2011), suggesting the conclusions were based largely on the fact that the genomes of distant teleost species retained two copies of core family members. Other phylogenetic studies concluding 3R have focused on single or a limited number of core family members with a relatively small number of sequences (Zhou et al. 2008; Dai et al. 2010; Shimizu, Kishimoto, et al. 2011; Shimizu, Suzuki, et al. 2011). Our study was based on robust sequence alignment/tree building methods and the most comprehensive representation of teleost IGFBP sequences to date, while being supported by gene family-wide synteny data. Consequently, the results allow numerous previous studies characterizing a limited number of teleost *IGFBP* genes to be placed within a common evolutionary context, facilitating interspecific comparisons of inferred gene function in light of 3R and 4R.

Conclusion

Paralog retention after 3R/4R provided a statistical signal allowing us to infer previously unreported differences in functions among vertebrate-wide IGFBP family members. As far as we are aware, the concept that recent gene family expansion can provide exploitable statistical power to help understand the evolution of ancient genetic functions has received little attention in the literature. Considering that numerous genes have been retained as multiple copies after WGD or other forms of gene duplication, an approach like ours may have relevance to future research aiming to understand the functional evolution of many eukaryotic gene families.

Materials and Methods

Animal Experiments

All experimental procedures and husbandry practices involving animals were conducted in compliance with the Animals Scientific Procedures Act 1986 (Home Office Code of Practice. HMSO: London January 1997) in accordance with EU regulation (EC Directive 86/609/EEC) and approved by the Animal Ethics and Welfare Committee of the University of St Andrews.

One hundred presmolt Atlantic salmons were transferred from the Institute of Aquaculture (University of Stirling) to the Scottish Oceans Institute (University of St Andrews, UK) in August 2012. Fish were held in a closed circulating freshwater system within the same tank at 12 °C with a 12 h light:12 h dark photoperiod and satiation-fed commercial pellets. Following 2-months acclimatization, four fish were randomly caught and sacrificed according to UK Home Office guidelines. Their mean masses and fork lengths (FL) were 46.2 g (3.88 g standard deviation [SD]) and 170 mm (6 mm SD), respectively. Dissected samples of whole-brain, skin, head-kidney, heart-ventricle, gill-filament, lower-intestine, whole eye, fast-twitch myotomal muscle, liver, spleen, and stomach were flash-frozen in liquid N_2 and stored at −80 °C.

The fasting–refeeding experiment was performed at the Niall Bromage Freshwater Research Facility, Buckieburn (near Stirling, UK). Twenty-four fish were held in a single tank, gravity fed from a nearby reservoir at an ambient temperature (average 14.6 °C) and satiation-fed commercial pellets. After 2 weeks acclimatization, four fish were randomly sacrificed with mean masses and FLs of 7.9 g (0.9 g SD) and 8.9 cm (1.5 cm SD), respectively. Remaining fish were subjected to 72 h complete feed-restriction followed by 18 h ad libitum feeding (as for the acclimatization period), with four fish sampled per time point. Fasted fish had mean masses and FLs of 9.6 g (2.2 g SD) and 9.7 cm (0.7 cm SD), respectively, and the refed fish of 9.0 g (0.9 g SD) and 9.5 cm (0.3 cm SD), respectively. At each time point, the liver was dissected and stored in RNA later (Ambion).

Databases and Transcriptome Assemblies

We utilized nuclear-genome sequence assemblies from Atlantic salmon (NCBI BioProject 72713) and the following Ensembl (http://www.ensembl.org/, last accessed February 11, 2013) assemblies (versions bracketed): *Danio rerio* (v.Zv9), *Gasterosteus aculeatus* (v. BROADS1), *Oreochromis niloticus* (v.Orenil1.0), *Oryzius lapites* (vMEDAKA1), *Takifugu rubripes* (vFUGU4), *Tetraodon nigroviridis* (vTETRAODON8), and *Homo sapiens* (v.GRCh37). Sanger trace-chromatograms were manually examined via BLASTn screening of trace archives (www.ncbi.nlm.nih.gov/blast/tracemb.shtml, last accessed February 11, 2013).

Transcriptome assemblies were generated for Atlantic salmon and rainbow trout. Roche 454 pyrosequences were obtained from the NCBI Sequence Read Archive (SRA accession numbers: SRX118090 and SRX017741 for Atlantic salmon; SRX041526, SRX085156, DRX000493, SRX007396,

and SRX041532 for rainbow trout). Other assembled data included all sequences for each species contained in the NCBI EST (498,212 and 287,967 respective sequences for Atlantic salmon and rainbow trout) and nucleotide databases (mRNAs only: 16,727 and 140,528 respective sequences for Atlantic salmon and rainbow trout). Reads were assembled using Newbler v.2.5 (Roche, 454 Life Sciences), employing default settings. Combined isotigs and contigs generated by Newbler were used to make local BLAST databases in BioEdit (Hall 1999) v.7.0.9.1. Newbler assemblies will be provided by request to D.J.M. and associated statistics are provided in the supplementary material S4, Supplementary Material online.

Comparative Genomics

Orthologs of known salmonid *IGFBP* genes (accession numbers: JF920120, EF432856, EF432858, EF432860, HM536183, GU933436, GU933434, GU933428, and EF432864) were employed in BLASTn searches (complete coding sequences) against Atlantic salmon genome contigs via NCBI genomic BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi, last accessed February 11, 2013) and against Atlantic salmon and rainbow trout transcriptome assemblies. This approach identified several *IGFBP* sequences sharing similarity consistent with 4R (>80% nucleotide identity). Corresponding complete coding-sequences were acquired manually, facilitated by conservation of coding/splicing features among putative paralogs. Intron–exon structures were determined using Spidey (http://www.ncbi.nlm.nih.gov/spidey/, last accessed February 11, 2013), aligning experimentally validated *IGFBP* mRNAs with *IGFBP*-containing contigs sharing the highest sequence similarity.

Before this study, single gene copies of *IGFBP-3* and *IGFBP-6* were characterized in salmonids, despite the presence of 3R paralog-pairs in other teleosts. In a successful attempt to identify salmonid orthologs of these genes, IGFBP-3 and -6 protein sequences from *G. aculeatus*, *O. latipes*, and *T. rubripes* were used in tBLASTn searches against salmon genome contigs. Although the contigs containing the genes of interest were identified, this approach could not reliably identify start and stop regions. Thus, genomic contigs were also submitted to Augustus (Stanke and Morgenstern 2005) to generate gene models. These data combined with tBLASTn results facilitated prediction of regions containing start and stop codons.

We also screened the Atlantic salmon genome and our salmonid transcriptome assemblies for unknown 3R or 4R copies of IGF1 and IGF2 using BLASTn searches with published Atlantic salmon sequences (respective accession numbers: AAA18211 and EF432854).

Synteny surrounding *IGFBP* genes was manually inferred by study of Ensembl assemblies.

Sanger Sequencing

Total RNA extraction, quality analysis, and concentration determination protocols are described elsewhere (Macqueen, Kristjánsson, et al. 2010; Macqueen et al. 2011). 10,000 ng of total RNA equally representing 11 Atlantic salmon tissues by concentration (from one individual described earlier) was reverse transcribed using QuantiTect reverse transcriptase (Qiagen) following the manufacturer's instructions, including for genomic DNA removal. 200 µl first-strand (FS)-cDNA was column purified (QIAquick spin column, Qiagen) and eluted in 50 µl nuclease-free water. 1 µl of FS-cDNA (200 ng reverse transcribed RNA) was used in standard reverse transcriptase-PCR (RT-PCR) reactions containing 400 nM sense and antisense strand primers designed to amplify targeted complete coding-sequences (supplementary material S3, Excel Sheet 5, Supplementary Material online). The polymerase was BIOTAQ (Bioline), buffered to the manufacturer's instructions. Cycling conditions included 1 cycle of 10 min at 95 °C, 35 cycles of 30 s at 95 °C, 30 s at 58 °C, 1 min at 72 °C and 1 cycle of 10 min at 72 °C. RT-PCR mixes were separated by agarose gel electrophoresis and double-stranded (DS)-cDNAs of the anticipated size extracted then column purified/eluted in 30 µl as described earlier. In certain cases, DS-cDNAs were used as templates (1 µl used) in a second round of RT-PCR, performed as described earlier with 20 additional cycles.

PCR products were ligated into pDrive cloning Vector (Qiagen) following the manufacturer's protocol and transformed into One Shot® TOP10 chemically competent *Escherichia coli* (Invitrogen), cultured on selective agar plates. For each cloned product, 15 colonies were picked into standard PCR mix containing a primer pair specific to the pDrive vector, which amplifies the insert and a small flanking sequence. Products were sequenced in sense and antisense orientations using BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems) using custom primers orientated 5′ to those used for the previous PCR. Sequences were read by an Applied Biosystems 3730 DNA sequencer (outsourced to Source BioScience LifeSciences, UK).

Phylogenetic Analysis

Sequence alignment was performed using PRANK (Löytynoja and Goldman 2008) through the GUIDANCE server (Penn et al. 2010) employing the GUIDANCE algorithm to assess alignment quality and filter sites below a confidence cut-off score of 0.95. Finished alignments (supplementary material S4, Supplementary Material online) were loaded into MEGA5 (Tamura et al. 2011) to establish the best-fitting amino-acid substitution models by ML. Bayesian information criterion statistics indicated JTT+G to be overwhelmingly best-fitting for all alignments. Tree-building was performed in MEGA5 using ML and NJ. ML was performed with the best fitting evolutionary model (i.e., JTT, with concurrent estimation of the among-site rate distribution parameter, α) and NJ with the JTT model and α fixed as per the ML estimate. By all approaches, nonparametric bootstrapping (1,000 iterations) provided branch support values.

Protein Complex Modeling

NBP and CBP sequences of Atlantic salmon IGFBPs were inferred using PROSITE (Sigrist et al. 2002). For all 19 proteins, NBP and CBP were submitted along with mature salmon IGF1 and IGF2 (respective accession numbers AAA18211 and EF432854) to the Protinfo PPC webserver (Kittichotirat

et al. 2009) and a PDB file of the template (RCSB accession: 2DSR). PDB files for all models are available on request to D.J.M. Model quality was assessed using tools available through the SWISS-MODEL webserver (Arnold et al. 2006). Global quality metrics inferred by QMEAN (Benkert et al. 2011) indicated overall model qualities to be comparable with the template (supplementary material S3, Excel Sheet 3, Supplementary Material online). Specifically, the QMEAN6 score for 2DSR was not different to the mean of the 19 models ($P = 0.695$ and 0.622, respectively, for IGF1 and IGF2 models; one way ANOVA). Local (per-residue) QMEAN score functions (Benkert et al. 2009) were also considered. Although IGFBP-1, -2, -3, -4, and -5 models were inferred to have comparable local quality with the template in all regions, inaccuracy was identified in the NBP thumb of all IGFBP-6 complexes. This probably reflects the presence of extended N-terminal sequences in IGFBP-6 sequences compared with the template, meaning the thumb regions were ab initio modeled (Kittichotirat et al. 2009). When the IGFBP-6 models were rendered, the most N-terminal residues were not located adjacent to IGF1R binding residues of IGF, which was in contrast to the template and other family member models. The absence of a modeled thumb, considered to be a common feature of all IGFBPs (Sitar et al. 2006), dramatically changed or ablated interactions between IGFBP-6 proteins and the hormone. Considering that the thumb region is critical to our study conclusions, we were left with no option but to accept that the IGFBP-6 thumb region could not be modeled accurately.

Models were rendered in UCSF Chimera v.1.6.1. (Pettersen et al. 2004) and atomic-level contacts inferred using the Find Clashes/Contacts tool, with van der Waals criteria optimized toward favorable contacts.

Protocol for qPCR

The qPCR experiments conformed to MIQE guidelines (Bustin et al. 2009). All pipetting was performed out of 96-well plates using a Research pro electronic multi-channel pipette (Eppendorf). RNA used for FS-cDNA synthesis showed perfect integrity and had 260/280 and 260/230 nm absorbance spectra of 1.9–2.2 and >2.2, respectively. 1,000 ng of total RNA extracted from 11 tissues of four fish (described earlier) and from the 12 fish used in the fasting–refeeding experiment was reverse transcribed as detailed earlier. FS-cDNAs were diluted either 100- or 20-fold in nuclease-free water. Minus-reverse transcriptase (−RT) controls were separately made for the tissue distribution and fasting/refeeding experiments. Each −RT reaction contained 1,000 ng total RNA (a pool equally representing all the samples used in each experiment by concentration) and all components of the cDNA synthesis with water replacing RT.

Primer pairs were designed to the 19 Atlantic salmon *IGFBP* genes, such that each primer would bind the most distinguishing available regions between 3R/4R paralogs (particularly at each primer's 3′) in exons separated by at least one intron (supplementary material S3, Excel Sheet 5, Supplementary Material online). 2R family members share negligible nucleotide sequence identity, so were not considered as a possible source of cross-amplification during primer design. Six primer pairs targeting candidate reference genes have been previously validated (Bower et al. 2008; Macqueen, Bower, et al. 2010; Macqueen et al. 2011).

15 µl qPCR reactions contained 6 µl of FS-cDNA, 7.5 µl SensiFAST SYBR Lo-ROX 2X master mix (Bioline) and 400 nM sense/antisense primers. Reactions were performed using an Mx30005P thermocycler (Stratagene), with 1 cycle of 2 min at 95 °C then 40 cycles of 10 s at 95 °C and 20 s at 65 °C, followed by dissociation analysis, where a single peak was observed in all cases. Each plate contained all samples in duplicate along with triplicate no template controls (NTCs, water in place of FS-cDNA) and triplicate −RT controls. Cq values were calculated from baseline-corrected ROX-normalized fluorescence data, with the threshold and baseline-range fixed across plates as 0.5 and 3–15 cycles, respectively. The only exception was the highly abundant 18S gene, where the baseline-range was set at 3–10 cycles. A cut-off of "no expression" was considered to represent four cycles below the lowest Cq in any NTC (generally 40 cycles; see supplementary material S3, Excel Sheets 5 and 6, Supplementary Material online). −RTs produced Cq values comparable with the NTC values.

Each qPCR assay's efficiency was calculated using LinRegPCRv.11 (Ruijter et al. 2009) following the author's recommendations. Cqs were exported to Genex v.4.4.2 (MultiD Analyses AB) and corrected for differences in efficiency before samples meeting the criteria of "no expression" were reset to 40 Cq. Normfinder (Andersen et al. 2004) was used to compare the suitability of *rps29*, *rps13*, *rpl4*, *18S*, *EF1-α*, and *hprt1* as reference genes with 1:100 cDNA dilutions from the tissue experiment. *Rps29* and *rps13* genes were most stably expressed globally and across-tissues and used as normalizing factors for both experiments. For some *IGFBP* genes, the tissue expression level was low, meaning the 1:20 FS-cDNAs were used to increase accuracy. Thus, *rps13* and *rps29* were also assayed with the 1:20 FS-cDNAs. Raw Cq data and normalized expression values are provided in the supplementary material S3, Excel Sheets 6 and 7, Supplementary Material online.

Statistics

Most statistics were performed in MINITAB v.13.2 (MINITAB Inc.). One-way ANOVA was used to establish variation in IGFBP–IGF complex contact data (supplementary material S3, Excel Sheets 1–3, Supplementary Material online), employing Fisher's test to identify significantly different family members with the individual error rate set at 0.01. Expression data used for statistics were ranked to ensure homoscedasticity and normality in the data residuals. In the tissue-distribution experiment, this was mainly required due to the massively higher *IGBFP* expression in liver than other tissues. In the fasting–refeeding experiment, this was required due to the enormous range of fold-regulation observed between nutritional states. Stepwise regression modeling employed an alpha-level of 0.05 for entry and removal of predictors of *IGFBP* gene expression. Mallows' Cp (Mallows

1973) was used to assess model fit. Spearman's correlation was used to compare the association in the combined or separate number of α and β contacts with ranks of fold-regulation observed between control versus fasted and fasted versus refed nutritional states. Spearman's correlation was used to compare the expression levels of *IGFBP* gene-pairs in tissues (supplementary material S3, Excel Sheet 4, Supplementary Material online) comparing ranks of 44 samples (11 tissues of 4 fish). Unsupervised hierarchical clustering was performed with the same ranked expression data using pvclust within R (The R Foundation for Statistical Computing, http://www.r-project.org/foundation/, last accessed February 11, 2013) (Suzuki and Shimodaira 2006) employing average-linkage and a dissimilarity-matrix based on correlation. 5,000 bootstrap iterations were used to generate approximately unbiased probability (AUP) cluster support values (Suzuki and Shimodaira 2006).

Supplementary Material

Supplementary materials S1–S4 and figures S1–S10 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Marine Alliance for Science and Technology for Scotland pooling initiative and the Scottish Funding Council grant number HR09011 and contributing institutions. Additional funding was received from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 222719—LIFECYCLE. The authors are indebted to Prof. Herve Migaud and Dr John Taylor (both from the University of Stirling) for performing the feeding experiment and providing all Atlantic salmon used in the study. The authors thank Al Reeve (University of St Andrews) for help with R (the R Foundation for Statistical Computing).

References

Allendorf FW, Thorgaard GH. 1984. Tetraploidy and the evolution of salmonid fishes. In: Turner BJ, editor. Evolutionary genetics of fishes. New York: Plenum Press. p. 1–53.

Andersen CL, Jensen JL, Ørntoft TF. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variation estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64: 5245–5250.

Andress DL, Birnbaum RS. 1992. Human osteoblast-derived insulin-like growth factor (IGF) binding protein-5 stimulates osteoblast mitogenesis and potentiates IGF action. *J Biol Chem.* 267:22467–22472.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* 22:195–201.

Benkert P, Biasini M, Schwede T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27:343–350.

Benkert P, Künzli M, Schwede T. 2009. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37:W510–W514.

Boisclair YR, Rhoads RP, Ueki I, Wang J, Ooi GT. 2001. The acid-labile subunit (ALS) of the 150 kDa IGF-binding protein complex: an important but forgotten component of the circulating IGF system. *J Endocrinol.* 170:63–70.

Bower NI, Johnston IA. 2010. Transcriptional regulation of the IGF signaling pathway by amino acids and insulin-like growth factors during myogenesis in Atlantic salmon. *PLoS One* 5:e11100.

Bower NI, Li X, Taylor R, Johnston IA. 2008. Switching to fast growth: the insulin-like growth factor (IGF) system in skeletal muscle of Atlantic salmon. *J Exp Biol.* 211:3859–3870.

Bustin SA, Benes V, Garson JA, et al. (12 co-authors). 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 55:611–622.

Clemmons DR. 2001. Use of mutagenesis to probe IGF-binding protein structure/function relationships. *Endocr Rev.* 22:800–817.

Clemmons DR, Underwood LE. 1991. Nutritional regulation of IGF-I and IGF binding proteins. *Ann Rev Nutr.* 11:393–412.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9:938–950.

Conover CA. 1992. Potentiation of insulin-like growth factor (IGF) action by IGF-binding protein-3: studies of underlying mechanism. *Endocrinology* 130:3191–3199.

Dai W, Kamei H, Zhao Y, Ding J, Du Z, Duan C. 2010. Duplicated zebrafish insulin-like growth factor binding protein-5 genes with split functional domains: evidence for evolutionarily conserved IGF binding, nuclear localization, and transactivation activity. *FASEB J.* 24:2020–2029.

Davidson WS, Koop BF, Jones SJ, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, Omholt SW. 2010. Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol.* 11:403.

Degger B, Upton Z, Soole K, Collet C, Richardson N. 2000. Comparison of recombinant barramundi and human insulin-like growth factor (IGF)-I in juvenile barramundi (*Lates calcarifer*): in vivo metabolic effects, association with circulating IGF-binding proteins, and tissue localisation. *Gen Comp Endocrinol.* 117:395–403.

Denley A, Cosgrove LJ, Booker GW, Wallace JC, Forbes BE. 2005. Molecular interactions of the IGF system. *Cytokine Growth Factor Rev.* 16:421–439.

Duan C, Clemmons DR. 1998. Differential expression and biological effects of insulin-like growth factor-binding protein-4 and -5 in vascular smooth muscle cells. *J Biol Chem.* 273:16836–16842.

Duan C, Ding J, Li Q, Tsai W, Pozios K. 1999. Insulin-like growth factor binding protein 2 is a growth inhibitory protein conserved in zebrafish. *Proc Natl Acad Sci U S A.* 96:15274–15279.

Duan C, Xu Q. 2005. Roles of insulin-like growth factor (IGF) binding proteins in regulating IGF actions. *Gen Comp Endocrinol.* 142:44–52.

Ewton DZ, Coolican SA, Mohan S, Chernausek SD, Florini JR. 1998. Modulation of insulin-like growth factor actions in L6A1 myoblasts by insulin-like growth factor binding protein (IGFBP)-4 and IGFBP-5: a dual role for IGFBP-5. *J Cell Physiol.* 177:47–57.

Firth SM, Baxter RC. 2002. Cellular actions of the insulin-like growth factor binding proteins. *Endocr Rev.* 23:824–854.

Garcia de la serrana D, Vieira VLA, Andree KB, Darias M, Estévez A, Gisbert E, Johnston IA. 2012. Development temperature has persistent effects on muscle growth responses in gilthead sea bream. *PLoS One* 7:e51884.

Gockerman A, Prevette T, Jones JI, Clemmons DR. 1995. Insulin-like growth factor (IGF)-binding proteins inhibit the smooth muscle cell migration responses to IGF-I and IGF-II. *Endocrinology* 136: 4168–4173.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.

Jaillon O, Aury JM, Brunet F, et al. (61 co-authors). 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.

Johnston IA, Bower NI, Macqueen DJ. 2011. Growth and the regulation of myotomal muscle mass in teleost fish. *J Exp Biol.* 214: 1617–1628.

Jones JI, Clemmons DR. 1995. Insulin-like growth factors and their binding proteins: biological actions. *Endocr Rev.* 16:3–34.

Kajimura S, Aida K, Duan C. 2005. Insulin-like growth factor-binding protein-1 (IGFBP-1) mediates hypoxia-induced embryonic growth

and developmental retardation. *Proc Natl Acad Sci U S A.* 102: 1240–1245.
Kalus W, Zweckstetter M, Renner C, et al. (11 co-authors). 1998. Structure of the IGF-binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5): implications for IGF and IGF-I receptor interactions. *EMBO J.* 17:6558–6572.
Kamangar BB, Gabillard JC, Bobe J. 2006. Insulin-like growth factor-binding protein (IGFBP)-1, -2, -3, -4, -5, and -6 and IGFBP-related protein 1 during rainbow trout postvitellogenesis and oocyte maturation: molecular characterization, expression profiles, and hormonal regulation. *Endocrinology* 147:2399–2410.
Kamei H, Lu L, Jiao S, Li Y, Gyrup C, Laursen LS, Oxvig C, Zhou J, Duan C. 2008. Duplication and diversification of the hypoxia-inducible IGFBP-1 gene in zebrafish. *PLoS One* 3:e3091.
Kavsan VM, Grebenjuk VA, Koval AP, Skorokhod AS, Roberts CT Jr, Leroith D. 1994. Isolation of a second nonallelic insulin-like growth factor I gene from the salmon genome. *DNA Cell Biol.* 13: 555–559.
Kiepe D, Ciarmatori S, Haarmann A, Tönshoff B. 2006. Differential expression of IGF system components in proliferating vs. differentiating growth plate chondrocytes: the functional role of IGFBP-5. *Am J Physiol Endocrinol Metab.* 290:E363–E371.
Kiepe D, Ulinski T, Powell DR, Durham SK, Mehls O, Tönshoff B. 2002. Differential effects of insulin-like growth factor binding proteins-1, -2, -3, and -6 on cultured growth plate chondrocytes. *Kidney Int.* 62: 591–600.
Kittichotirat W, Guerquin M, Bumgarner RE, Samudrala R. 2009. Protinfo PPC: a web server for atomic level prediction of protein complexes. *Nucleic Acids Res.* 37:W519–W525.
Koop BF, Davidson WS. 2008. Genomics and the genome duplication in salmonids. In: Tsukamoto K, Kawamura R, Takeuchi T, Beard TD Jr, Kaiser MJ, editors. Fisheries for global welfare and environment, 5th World Fisheries Congress. Tokyo (Japan): Terrapub. p. 77–86.
Lee PD, Giudice LC, Conover CA, Powell DR. 1997. Insulin-like growth factor binding protein-1: recent findings and new directions. *Proc Soc Exp Biol Med.* 216:319–357.
Lemozy S, Pucilowska JB, Underwood LE. 1994. Reduction of insulin-like growth factor-I (IGF-I) in protein-restricted rats is associated with differential regulation of IGF-binding protein messenger ribonucleic acids in liver and kidney, and peptides in liver and serum. *Endocrinology* 135:617–623.
Leong JS, Jantzen SG, von Schalburg KR, et al. (12 co-authors). 2010. *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics* 11:279.
Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320:1632–1635.
Macqueen DJ, Bower NI, Johnston IA. 2010. Positioning the expanded akirin gene family of Atlantic salmon within the transcriptional networks of myogenesis. *Biochem Biophys Res Commun.* 400: 599–605.
Macqueen DJ, Delbridge ML, Manthri S, Johnston IA. 2010. A newly classified vertebrate calpain protease, directly ancestral to CAPN1 and 2, episodically evolved a restricted physiological function in placental mammals. *Mol Biol Evol.* 27:1886–1902.
Macqueen DJ, Kristjánsson BK, Johnston IA. 2010. Salmonid genomes have a remarkably expanded akirin family, coexpressed with genes from conserved pathways governing skeletal muscle growth and catabolism. *Physiol Genomics* 42:134–148.
Macqueen DJ, Kristjánsson BK, Paxton CG, Vieira VL, Johnston IA. 2011. The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. *Mol Ecol.* 20:3167–3184.
Mallows CL. 1973. Some comments on CP. *Technometrics* 15:661–675.
Mungpakdee S, Seo HC, Angotzi AR, Dong X, Akalin A, Chourrout D. 2008. Differential evolution of the 13 Atlantic salmon Hox clusters. *Mol Biol Evol.* 25:1333–1343.
Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci U S A.* 109:13698–13703.
Ocampo Daza D, Sundström G, Bergqvist CA, Duan C, Larhammar D. 2011. Evolution of the insulin-like growth factor binding protein (IGFBP) family. *Endocrinology* 152:2278–2289.
Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.
Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* 131: 452–462.
Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet.* 34:401–437.
Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38:W23–W28.
Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 25: 1605–1612.
Phillips LS, Pao CI, Villafuerte BC. 1998. Molecular regulation of insulin-like growth factor-I and its principal binding protein, IGFBP-3. *Prog Nucleic Acid Res Mol Biol.* 60:195–265.
Power DM, Melo J, Santos CRA. 2000. The effect of food deprivation and refeeding on the liver, thyroid hormones and transthyretin in sea bream. *J Fish Biol.* 56:374–387.
Putnam NH, Butts T, Ferrier DE, et al. (37 co-authors). 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
Ramagnolo D, Akers RM, Byatt JC, Wong EA, Turner JD. 1994. IGF-I-induced IGFBP-3 potentiates the mitogenic actions of IGF-I in mammary epithelial MD-IGF-I cells. *Mol Cell Endocrinol.* 102: 131–139.
Ren H, Yin P, Duan C. 2008. IGFBP-5 regulates muscle cell differentiation by binding to IGF-II and switching on the IGF-II auto-regulation loop. *J Cell Biol.* 182:979–991.
Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, van den Hoff MJ, Moorman AF. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* 37:e45.
Schneider MR, Wolf E, Hoeflich A, Lahm H. 2002. IGF-binding protein-5: flexible player in the IGF system and effector on its own. *J Endocrinol.* 172:423–440.
Shimizu M, Beckman BR, Hara A, Dickhoff WW. 2006. Measurement of circulating salmon IGF binding protein-1: assay development, response to feeding ration and temperature, and relation to growth parameters. *J Endocrinol.* 188:101–110.
Shimizu M, Kishimoto K, Yamaguchi T, Nakano Y, Hara A, Dickhoff WW. 2011. Circulating salmon 28- and 22-kDa insulin-like growth factor binding proteins (IGFBPs) are co-orthologs of IGFBP-1. *Gen Comp Endocrinol.* 174:97–106.
Shimizu M, Suzuki S, Horikoshi M, Hara A, Dickhoff WW. 2011. Circulating salmon 41-kDa insulin-like growth factor binding protein (IGFBP) is not IGFBP-3 but an IGFBP-2 subtype. *Gen Comp Endocrinol.* 171:326–331.
Shimizu M, Swanson P, Dickhoff WW. 1999. Free and protein-bound insulin-like growth factor-I (IGF-I) and IGF-binding proteins in plasma of coho salmon, *Oncorhynchus kisutch*. *Gen Comp Endocrinol.* 115:398–405.
Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform.* 3:265–274.
Sitar T, Popowicz GM, Siwanowicz I, Huber R, Holak TA. 2006. Structural basis for the inhibition of insulin-like growth factors by insulin-like growth factor-binding proteins. *Proc Natl Acad Sci U S A.* 103: 13028–13033.
Siwanowicz I, Popowicz GM, Wisniewska M, Huber R, Kuenkele KP, Lang K, Engh RA, Holak TA. 2005. Structural basis for the regulation of insulin-like growth factors by IGF binding proteins. *Structure* 13: 155–167.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33:W465–W467.

Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22: 1540–1542.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.

Taylor JS, Rae J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Ann Rev Genet.* 38:615–643.

Tseng LY, Ooi GT, Brown AL, Straus DS, Rechler MM. 1992. Transcription of the insulin-like growth factor-binding protein-2 gene is increased in neonatal and fasted adult rat liver. *Mol Endocrinol.* 6: 1195–1201.

Underwood LE. 1996. Nutritional regulation of IGF-I and IGFBPs. *J Pediatr Endocrinol Metab.* 9(Suppl 3):303–312.

Upton Z, Chan SJ, Steiner DF, Wallace JC, Ballard FJ. 1993. Evolution of insulin-like growth factor binding proteins. *Growth Regul.* 3:29–32.

Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplication. *Nat Rev Genet.* 10:725–732.

Wallis AE, Devlin RH. 1993. Duplicate insulin-like growth factor-I genes in salmon display alternative splicing pathways. *Mol Endocrinol.* 7: 409–422.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet.* 2:333–341.

Yamada PM, Lee KW. 2009. Perspectives in mammalian IGFBP-3 biology: local vs. systemic action. *Am J Physiol Cell Physiol.* 296: C954–C976.

Zhou J, Li W, Kamei H, Duan C. 2008. Duplication of the IGFBP-2 gene in teleost fish: protein structure and functionality conservation and gene expression divergence. *PLoS One* 3:e3926.

Zhu W, Shiojima I, Ito Y, et al. (15 co-authors). 2008. IGFBP-4 is an inhibitor of canonical Wnt signalling required for cardiogenesis. *Nature* 454:345–349.